Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Air Force Personnel and Training Research Center

DAVID A. GRANT, Associate Editor
University of Wisconsin

DELOS D. WICKENS, Associate Editor
Ohio State University

Consulting Editors

E. JAMES ARCHER, *University of Wisconsin*
JUDSON S. BROWN, *State University of Iowa*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *Ohio State University*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *AF Personnel and Training Research Center*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
NEAL E. MILLER, *Yale University*
KENNETH W. SPENCE, *State University of Iowa*

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Assistant Managing Editor
Editorial Staff: SADIE J. DOYLE, FRANCES H. CLARK, BARBARA CUMMINGS

CONTENTS

Effects of Course Frequency and Aided Time Constant on Pursuit and Compensatory Tracking: R. CHERNIKOFF AND F. V. TAYLOR 285

Proactive Interference and Facilitation as a Function of Amount of Training and Stress: D. S. PALERMO 293

Serial Effects in Auditory Threshold Judgments: W. J. MCGILL 297

The Abstraction and Conceptualization of Form, Color, and Number: J. F. WOHLWILL ... 304

The Effect of Drive Reversal on Latency, Amplitude, and Activity Level: R. H. DAVIS 310

Response Strength as a Function of Delay of Reward in a Runway: W. B. HOLDER, M. H. MARX, E. E. HOLDER, AND G. COLLIER 316

Factors Contributing to General Versus Specific Perceptual Learning: J. A. LEONARD, H. W. CLARKE, AND S. R. STAATS 324

Application of the Half-Split Technique to Problem-Solving Tasks: R. A. GOLDBECK, B. B. BERNSTEIN, W. A. HILLIX, AND M. H. MARX 330

The Relation of Eye Movements During Sleep to Dream Activity: An Objective Method for the Study of Dreaming: W. DEMENT AND N. KLEITMAN 339

Effect of Irrelevant Thirst Motivation on a Response Learned with Food Reward: G. R. GRICE AND J. D. DAVIS 347

Prediction From and Interaction Among Multiple Concurrent Discriminative Responses: C. W. ERIKSEN 353

Supplementary Report: Response Availability in Human Trial-and-Error Learning: C. E. NOBLE 359

American Psychological Association

Vol. 53 No. 5 May 1957

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal, experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research effort is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this JOURNAL, 1957, 53, 1–2).

Address all articles submitted for publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, P. O. Box 1557, Lackland Air Force Base, Texas. All manuscripts should be submitted in two complete copies; only the original copy will be returned if the article is not accepted for publication. Figures may be original line drawings or glossy photographic prints, but should not exceed 8½ by 11 in. Care should be taken to insure legibility of labels and critical features of drawings when reduced to a single-column or (exceptionally) a double-column figure in the JOURNAL.

To save printing costs supplementary material to articles in this JOURNAL are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 [4, Suppl.], 389–449). In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1957 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 53, NO. 5 MAY, 1957

EFFECTS OF COURSE FREQUENCY AND AIDED TIME CONSTANT ON PURSUIT AND COMPENSATORY TRACKING

RUBE CHERNIKOFF AND FRANKLIN V. TAYLOR[1]

Naval Research Laboratory

Previous studies (**2, 3, 6, 9**) have shown that independent variations in display mode (pursuit vs. compensatory), aided tracking time constant, and frequency of the target course input have important effects upon tracking performance. The purpose of this experiment is to study the interaction among these parameters.

In pursuit tracking, *S* is required to position a control marker to coincide with an independently driven target course marker. With compensatory tracking, *S*'s task is to keep a moving marker which is positioned by the difference between the target input and the control output, aligned with a reference marker fixed at the center of the display.

The simplest form of aided tracking obtains when a movement of the control causes a proportional change in the position of the control marker, and simultaneously introduces a change in marker velocity as well. The ratio of the position sensitivity to the velocity sensitivity is termed the aiding time constant. This constant is computed as follows: Aiding time constant (in sec.) = marker displacement per degree of control displacement (in inches per degree) divided by marker velocity per degree of control displacement (in inches per sec. per degree).

In a Foxboro Company report (**4**) it was first suggested that aided tracking be considered the general case, with position tracking one extreme special case containing no velocity component, and velocity (or rate) tracking the other extreme, in which no position component is present. By thus considering tracking control systems on an aiding continuum, position control is defined as having an aiding time constant of infinity (∞), and rate control as having a time constant of zero. (The value of all time constants in this report is in seconds.) These two limiting time constants were employed in this experiment along with an intermediate value of .5 sec. The latter was chosen because it is frequently taken by engineers to represent an optimum value for practical gun fire control systems. Tracking

[1] The authors are indebted to Mr. Henry P. Birmingham for his assistance in the equipment design and for the many suggestions made during the formulation of the problem.

performance with these three time constants for both pursuit and compensatory modes was compared on three different target courses, composed of frequencies chosen to cover a wide range of tracking difficulty.

Method

Apparatus

Display.—A 5-in. cathode ray tube (CRT) with a P-11 short-persistence screen was used to display the tracking information. By means of an electronic switch, the CRT beam was time-shared to produce two spots of light. One spot was shaped into a ½-in. vertical line, while the other was left as a $\frac{1}{16}$-in. diameter dot. During pursuit tracking, the vertical line was controlled by *S* and the dot positioned by a course generator. When compensatory tracking was used, the dot was fixed in the center of the CRT face as the reference, while the vertical line was driven by the difference between the course generator output and *S*'s control movements.

Control.—The *S* positioned the vertical line marker by left-right movements of a horizontally mounted joy stick. This stick consisted of a flexible steel bar, ¼ in. thick, ½ in. wide, and 24 in. long, with a Statham strain gauge attached to the bar to provide a noise-free pick-off of control stick motion. With position tracking (∞), the joy stick-to-marker sensitivity was set so that a lateral stick movement of ½ in. (by a force application of 4 oz.) resulted in a 1-in. displacement of the marker on the CRT face. (Stick displacement is given in inches because of the difficulty in specifying the angular displacement of the flexible control.) With rate tracking (0) a stick displacement of ½ in. produced a marker velocity of 2 in./sec. With .5 aiding the relationship between the joy stick and marker was such that a ½-in. stick displacement produced both a position change of 1 in. and a velocity change of 2 in./sec.

Courses.—Three courses, each consisting of a complex of three sine waves, were used in the experiments. The frequencies comprising each of the courses are listed in Table 1. For all courses, the relative amplitude of each sine wave was inversely proportional to its frequency. This provided an equal maximum rate for each frequency of the course. As displayed, all three courses were of the same amplitude and the maximum excursion of the marker was across the center 4 in. of the CRT face.

Scoring.—Since an AC voltage, which is proportional in amplitude to the separation between the spots on the CRT face, exists at the output of the electronic switch, it was possible to obtain directly a measure of spot separation. By use of an electronic integrator, this error voltage was rectified and then integrated over the scoring duration by cumulating a charge on a condenser. Integrated error was read directly on a voltmeter.

TABLE 1

Sine Wave Frequencies Comprising Each of the Courses

Course	Frequencies (Cycles per Minute)		
A	6 2/3	4 4/9	2 2/3
B	16 2/3	11 1/9	6 2/3
C	26 2/3	17 7/9	10 2/3

Procedure

The *S* sat facing the CRT display at a viewing distance of approximately 12 in. A fluorescent desk lamp was placed behind the CRT to provide room illumination with a minimum of reflection from the CRT face. The joy stick was mounted 10 in. to the right of the CRT centerline and 2 in. above the height of *S*'s arm rest.

Before the experimental sessions were begun, two training sessions consisting of four practice trials per condition were given to familiarize *S* with the equipment. No attempt was made to train *S* to respond in any specific manner to the various conditions. For pursuit tracking, *S* was instructed to move the joy stick to the left or right so as to keep the vertical line marker on the moving course dot, while for compensatory he was instructed to keep the vertical line marker on the fixed reference dot.

Three experiments were run utilizing the same design but differing in the course frequencies which comprised the tracking problem. The order in which the courses were run was as follows: Exp. I, Course B; Exp. II, Course A; and Exp. III, Course C. Eighteen Navy enlisted men served as *S*s, six in each experiment. The six conditions, ∞-compensatory (∞–C), 0-compensatory (0–C), .5-compensatory (.5–C), ∞-pursuit (∞–P), 0-pursuit (0–P), and .5-pursuit (.5–P), were presented in a balanced latin-square design in blocks of three trials per condition for 18 daily sessions. One practice trial always preceded each block of scored trials. A trial was 1 min. in duration, with the last 45 sec. scored by the error integrator. To reduce fatigue effects from an overly long tracking period, the six conditions were presented in two sittings of three conditions each. The interval between successive sittings was approximately 4 hr.

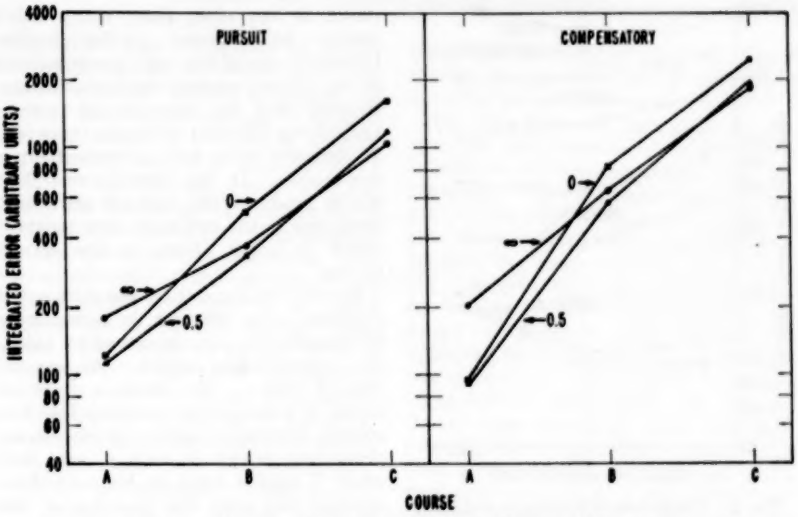


FIG. 1. Tracking performance data showing interaction between aiding time constants and course frequency.

The 1-min. course duration and 45-sec. scoring interval were automatically programmed. After each trial, *E* recorded the integrated-error score and allowed a 30-sec. rest interval before starting the next trial.

RESULTS AND DISCUSSION

Since the main interest in this study was comparative tracking performance after S had reached a fairly stable level, the statistical analyses and graphical representations are based on the data collected during the last four days of each experiment (Sessions 15–18). Figure 1 presents a graph of the tracking performance data showing the interaction between the aiding time constants and the course difficulty levels. Application of Bartlett's test to the data for each of the three experiments indicated a lack of homogeneity of variance for the measures of one of the experiments (Course A). Therefore, significance of differences between the various conditions was determined by using Wilcoxon's (**10**) ranking procedure for paired replicates. Comparisons between conditions were based on the data for 6 *S*s, replicated over the last four sessions. The results of the nonparametric *T* tests are shown in Table 2.

In Fig. 2 the data are plotted in a way so as to afford a direct comparison of tracking performance on pursuit and compensatory displays. Three results are apparent from the data: (*a*) From Fig. 1, tracking error with

TABLE 2

SIGNIFICANCE LEVELS FOR COMPARISONS BETWEEN CONDITIONS USING NONPARAMETRIC *T* TESTS

Comparisons	Course A	Course B	Course C
0–P vs. .5–P	.01	.01	.01
0–P vs. ∞–P	.01	.01	.01
.5–P vs. ∞–P	.01	N.S.	.01
0–C vs. .5–C	N.S.	.01	.01
0–C vs. ∞–C	.01	.02	.01
.5–C vs. ∞–C	.01	N.S.	.05
0–P vs. 0–C	.01	.01	.01
.5–P vs. .5–C	.01	.01	.01
∞–P vs. ∞–C	.01	.01	.01

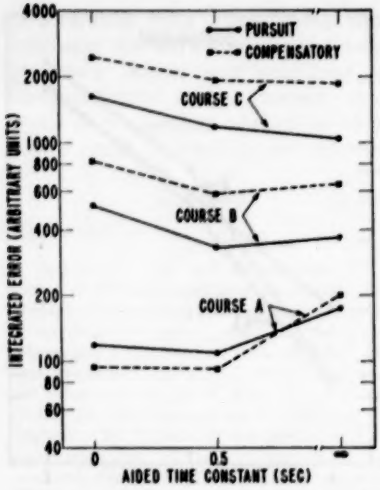


FIG. 2. Comparison of tracking performance on pursuit and compensatory displays as a function of aiding time constants and course frequency.

both display modes in association with all three time constants increases rapidly, with accompanying shifts in the relative proficiency of the various time constants, as the target course frequencies are increased. (*b*) Also from Fig. 1, with both pursuit and compensatory tracking, the optimum aided tracking time constant shifts upward with increasing target course frequencies. However, for the total range of frequencies covered, the best single time constant of those tested was .5. (*c*) From Fig. 2, pursuit tracking with a time constant of ∞ is superior to compensatory at all frequencies. The pursuit mode is also superior to the compensatory with time constants of 0 and .5 at the high and medium frequencies, but with these two time constants at the lowest frequencies, compensatory tracking becomes superior to pursuit.

The fact that tracking error increases as the frequencies in the target course are raised is not new, since many other studies have shown similar results. However, one of the main contributions of the present experiment is the demonstration that the accuracy of systems embodying different time constants falls at different rates with increased input frequencies. It is, indeed, this fact which produces the upward shift with frequency in the optimum time constant which is noted above as the second finding.

In order to account for the shift in the optimum time constant, it is necessary to consider the purpose served by aiding in a man-machine system. As was explained earlier, the simplest form of aiding is achieved by inserting into the control mechanism some type of velocity generating device in such a way that when *S* applies force to his control he changes not only the position of the controlled marker but also the velocity at which the marker is being driven. The velocity and the position terms play different roles in the aiding process.

The velocity component has several different effects, with the three most important being (*a*) the "unburdening" of *S* through making it unnecessary for him to track the velocities in the target course (1), (*b*) the introduction of a time lag into the control system, and (*c*) the increased complexity of *S*'s pattern of force application in making quick step function positional corrections. This factor is felt to be most important when high frequency courses are being tracked. Under such conditions *S*'s ability to follow continuously is taxed to the point where the tracking problem begins to resemble that posed by a course composed of a succession of step function changes. In regard to the matter of complexity, *S*, when using pure position control, needs only to apply a force and maintain it in order to move the controlled quantity from one stationary position to another. However, with velocity control, the same correction calls for an application of force followed by its removal.

In general, the unburdening effect of the velocity generating device assists *S*, while the increased complexity tends to

work against him. The time lag probably exerts two effects which act in opposite directions. On the one hand, the lag may benefit performance, at least at low frequencies, as a result of filtering out or damping *S*'s high-frequency "noise" (**8**). At the same time, it delays *S*'s knowledge of the results of his corrective actions and thus tends to degrade performance. However, this detrimental effect of the time delay is to a great extent overcome in an aided tracking device by the position component. For when tracking with a spring-restrained joy stick, the position term represents directly the force output of *S* and thereby serves to give him completely up-to-date visual information about what he is doing with his hand.

Aided tracking, then, provides a complex compromise of factors, some of which assist *S* by reducing his task while others act in an opposing fashion to make his task more difficult. It may be suggested that the shift in optimum aided tracking time constant observed in the experiment is to be accounted for by a change in the balance of these factors as the input frequencies are altered. Thus, at very low frequencies, where the major portion of the tracking task consists of matching the slowly changing velocity of the target, the beneficial aspects of the velocity term might be expected to predominate over its deleterious effects. This would suggest that the velocity term should assume a large value and hence a low time constant would be called for. On the other hand, with high target frequencies, the velocities are changing with great rapidity so that very little benefit can be anticipated from unburdening and damping, whereas the disruptive effects of the velocity component now exert their full force. This would mean that a high valued time constant would be required for the precise tracking of target courses in which high frequencies predominate.

Finally, with intermediate target frequencies, one would anticipate that the greatest benefit would be obtained by balancing the help furnished by each of the two terms through providing appreciable values of both. The aided tracking time constant of .5 sec. would thereby be expected to be superior to the other two values tested.

The direction of the shift in the optimum time constant observed in the experiment is consonant with this explanation. With the high-frequency target, ∞ was clearly the best, whereas .5 gave the lowest error scores at the intermediate frequency, although it did not differ significantly from ∞ at this point. At the low target frequency, where it would be expected that 0 would be optimum, the 0- and .5-sec. constants did not differ from one another with the compensatory mode, and with pursuit, .5 sec. was associated with the best tracking at this frequency. However, from a consideration of the slopes of the curves, it would appear that with lower frequencies than those employed in this experiment, the 0- and .5-sec. curves would have crossed as predicted.

A clearer way of noting the trend is to consider the data for the time constant of ∞. At the high end it is the best value, at the middle frequency it drops to a position intermediate between 0 and .5 (it is significantly different only from 0) and with the low-frequency target course it is the worst of the three constants.

The findings suggest that it would be not only desirable to extend the tests to a lower frequency than any utilized in the experiment, but that the medium course frequencies which were employed may have been too high. With a somewhat lower intermediate frequency, .5 sec. might then become significantly superior to the other two values.

The results lead to the practical suggestion that if control or tracking devices are to be built to work within fairly narrow frequency ranges, time constants should be inserted which are appropriate to the expected range, e.g., high constants for high frequencies and low constants for low frequencies. However, if it is known in advance that the system must perform acceptably over a very wide range of input frequencies, the single best constant to be recommended is one at or near .5 sec. Better still would be a method of appropriately shifting the aided tracking time constant as the input frequencies changed.

The third result of the experiment is that pursuit tracking is superior to compensatory at medium and high frequencies for all three time constants, but that the pursuit mode loses its advantage at the low frequencies when a velocity term is incorporated in the tracking system.

It is entirely likely that the general superiority of the pursuit tracking mode over the compensatory arises from the additional information presented to *S* by the pursuit display. The compensatory indication presents directly to *S* only visual error in terms of direction and amplitude. No immediate inference as to the nature of the target course is permitted since motions of the indicator due to the target input and movements due to *S*'s corrections are mutually contaminated through being mixed together. This contamination becomes more and more troublesome as the frequencies of the target are raised.

In contrast to this "single channel" indication, the pursuit display presents three channels of information rather than only one. Visual error is here shown as the distance between the target course marker and the indicator which represents the response of the system to *S*'s force applications. By means of this one display, information is provided to *S* which permits him to take account of the properties of the target and the tracking system independently of one another, as well as the nature of the error.

From these considerations, it is not difficult to account for the superiority of pursuit over compensatory tracking found for the majority of test points employed in the experiment. It is important to add, however, that this obtained result does not take account of the fact that the intrinsic properties of a compensatory display permit its sensitivity to be increased considerably, whereas this is not usually true of the pursuit mode. In the present experiment this potential advantage of compensatory tracking was not exploited, as it could be in engineering practice, and therefore the study does not furnish a direct comparison of the two modes with both fully optimized. In view of this, it is particularly striking to note that for the lowest target course frequencies, the superiority of the pursuit display disappears for aiding ratios of .5 and 0 and that compensatory tracking becomes the more accurate. Yet this finding is quite understandable when the effects of aiding discussed next are considered.

With low target frequencies it becomes possible, through the use of the velocity component, to so completely unburden *S* that his task becomes extremely simple. When thus unburdened, a close approach is made to what has been called "stimulus-response integrity" (1). Under this condition, the information contained in the instantaneous amplitude value of the visual error is sufficient to specify completely the response which will reduce this value to zero. Thus, if perfect stimulus-response integrity could be achieved, an invariant and proportional relationship would be found to obtain between error amplitude and the required *S* response. The *S* would not have to take account of target velocities or error velocities nor would he have to predict in any fashion.

Although perfect stimulus-response integrity cannot be achieved fully in practice, it can be approached so closely with low-frequency inputs that *S*, with his inherent perceptual and motor limitations, can do his best tracking only by attending fully to error amplitude and by rejecting all other information. It seems possible, therefore, to account for the superiority of compensatory tracking over pursuit at time constants of .5 and 0 with the lowest target frequencies on the basis of the fact that it is easier to give full attention to error on the compensatory than on the pursuit display.

The findings of this part of the study lead to the recommendation that a compensatory, rather than a pursuit, display should be utilized in practical control systems intended for following low-frequency inputs if a velocity term is involved. The marked superiority of pursuit tracking over compensatory at the higher frequencies might suggest

that, where possible, the pursuit mode should be employed in control devices intended for following high-frequency inputs. However, since no effort was made in the experiment to optimize the sensitivity of the compensatory display, the possibility is not precluded that compensatory tracking of high-frequency targets could be made to be much more precise than is shown in this study.

In view of the results of the present experiment, it may be suggested that some of the discrepancies in the tracking literature may have arisen as a result of various investigators employing target courses of different frequency ranges and distributions. Much of the early engineering work on aided tracking, which showed improvement in tracking accuracy with the introduction of the velocity term, was carried out with actual or simulated gun-fire control target courses which consist predominantly of low-frequency components. In contrast to this, the tracking research carried out in psychological laboratories has typically employed target courses of relatively high frequency, against which a velocity component in the tracking system is ineffective or actually disruptive.

Thus, the findings of Lincoln and Smith (**5**) that pure position tracking is superior to tracking with time constants of .5 and 0 sec., may be accounted for if it is assumed that they employed relatively high target frequencies comparable to our Course C. Although the frequency distribution of their course is not specified, the general description of their cam suggests that this might be the case.

In the present study, course difficulty was altered by changing the frequencies in the course while holding the amplitude constant. That a different result would have been obtained if the difficulty had been changed by altering the amplitude while holding frequency constant is indicated by an experiment by Pearl, Simon, and Smith (**7**). Working with a pursuit display, they compared tracking precision with time constants of .25, .5 and 1.0 sec. at three levels of course difficulty. These stages of difficulty were achieved by altering course speed and amplitude together while holding the frequency components of the course unchanged. Although their results are somewhat difficult to interpret, there is an indication in their data of a *downward* shift in the optimum time constant with an increase in course difficulty. This provides an interesting contrast to the present results wherein the optimum shifted *upward* with increased difficulty brought about by upping the target course frequencies.

A previous study (**3**) comparing pursuit and compensatory tracking with a control system including two integrators showed no difference between the two modes on a course containing frequencies lower than the present study's Course A, in which compensatory was superior with time constants of 0 and .5. This apparent discrepancy in results between the two studies at the low frequencies may have been due to differences in the design and conduct of the experiments, the types of tracking controls, the sensitivities of the scoring systems, or the number of integrators in the control systems. Further research may isolate the important variables determining the precision of pursuit and compensatory tracking at low course frequencies.

Summary

From previous studies, the independent variations in display mode (pursuit vs. compensatory), aided tracking time constant, and frequency of the target course input were found to affect the precision of tracking performance. This experiment was designed to study the interaction among these parameters.

Tracking performance with aiding time constants of 0, .5, and ∞ for both the pursuit and compensatory modes was compared on three target courses, composed of frequencies chosen to cover a wide range of tracking difficulty. Three studies were run, differing only in the course input used. In each study, *S*s were given three trials daily for 18 sessions, on all six combinations of time constants and display modes. A trial was 1 min. in duration, with the last 45 sec. scored by an error integrator. Each course was composed of three sine waves; the display was a 5-in. cathode ray tube; and the control was a spring-centered joy stick. The following results were obtained:

1. As the course frequencies increased, there was a rapid rise in tracking error, with accompanying shifts in the relative proficiency of the various time constants.

2. With both pursuit and compensatory tracking, the optimum aiding constant shifts upward with increasing course frequencies, although for the total range of frequencies covered the single best constant was .5.

3. Pursuit tracking is superior to compensatory with all time constants for the two courses of high and medium frequencies. However, the course composed of the lowest frequencies showed pursuit to be better than compensatory only at the time constant of ∞, whereas compensatory was superior using 0 and .5 aiding constants.

A theoretical analysis of the shifts in the proficiency of the aiding constants and the changes in effectiveness of the two display modes, as related to the change in course frequencies, is presented.

References

1. Birmingham, H. P., & Taylor, F. V. A human engineering approach to the design of man-operated continuous control systems. U. S. Naval Res. Lab. Rep. No. 4333, 1954.
2. Chernikoff, R., Birmingham, H. P., & Taylor, F. V. A comparison of pursuit and compensatory tracking under conditions of aiding and no aiding. *J. exp. Psychol.*, 1955, **49**, 55–59.
3. Chernikoff, R., Birmingham, H. P., & Taylor, F. V. A comparison of pursuit and compensatory tracking in a simulated aircraft control loop. *J. appl. Psychol.*, 1956, **40**, 47–52.
4. Foxboro Company. Studies in aided tracking. Memorandum No. 25 to Div. 7, N.D.R.C., Foxboro Co., Foxboro, Mass., 1945, 1–43.
5. Lincoln, R. S., & Smith, K. U. Systematic analysis of factors determining accuracy in visual tracking. *Science*, 1952, **116** (3008), 183–187.
6. Noble, M., Fitts, P. M., & Warren, C. E. The frequency response of skilled subjects in a pursuit tracking task. *J. exp. Psychol.*, 1955, **49**, 249–256.
7. Pearl, B. E., Simon, J. R., & Smith, K. U. Visual tracking: IV. Interrelations of target speed and aided-tracking ratio in defining tracking accuracy. *J. appl. Psychol.*, 1955, **39**, 209–214.
8. Rockway, M. R. The effect of variations in control-display ratio and exponential time delay on tracking performance. USAF, WADC Tech. Rep., 1954, No. 54-618.
9. Searle, L. V. Psychological studies of tracking behavior. Part IV—The intermittency hypothesis as a basis for predicting optimum aided-tracking time constants. U. S. Naval Res. Lab. Rep. No. 3872, 1951.
10. Wilcoxon, R. *Some rapid approximate statistical procedures*. Stamford, Conn.: American Cyanamid Co., 1949.

(Received May 14, 1956)

Journal of Experimental Psychology
Vol. 53, No. 5, 1957

PROACTIVE INTERFERENCE AND FACILITATION AS A FUNCTION OF AMOUNT OF TRAINING AND STRESS [1]

DAVID S. PALERMO [2]

Iowa Child Welfare Research Station, State University of Iowa

In a recent study, Castaneda (2) found a significant interaction between stress and type of S-R pair learned in a second task. One type of S-R pair was identical to those learned in a previous task, and the other type involved re-pairing of stimuli and responses learned in the first task, a procedure suggested by Duncan (4) and Porter and Duncan (6). The interaction was based on a tendency for performance of the stress group to be poorer than that of the nonstress group where re-pairing was required, but better where it was not. Assuming that the incorrect response tendency was dominant where re-pairing was required, while the correct response was dominant where it was not, and that the pacing condition (stress) served to increase the drive level, these results were interpreted as consistent with Spence's derivations (7) from Hull (5). Using a similar procedure, Castaneda and Palermo (3) obtained results suggesting that the magnitude of impairment or facilitation of performance as a result of an increase in drive varies as a positive function of the strength of the dominant habit.

The present experiment employed a modification of the procedure reported elsewhere by Castaneda (1) to investigate the effects of stress on performance of a task in which the dominant habit associated with a given stimulus was either a correct habit which was either relatively weak or strong, or an incorrect habit which was either relatively weak or strong.

METHOD

Subjects and design.—A total of 78 fourth-grade *Ss* participated in the present experiment.[3] However, following the procedure reported by Castaneda (2), 27 of these *Ss* were eliminated from the analysis because of failure to attain a criterion of 16 correct responses in the first task. Of the remaining 51 *Ss*, 25 were assigned to the paced (stress) group and 26 to the nonpaced (nonstress) group.

Apparatus and procedure.—The apparatus, identical to that reported by Castaneda (2), consisted of a box with a response panel on which was located two push buttons and a green jeweled reflector. Above the panel the four colored stimulus lights were presented through an aperture. If the correct button was selected, the stimulus light went off and the green light went on. Incorrect responses affected neither the stimulus light nor the light on the response panel.

The procedure was the same as reported by Castaneda (2) except that in the first task 20 trials were given such that two of the lights appeared eight times apiece, two appeared twice apiece, and no single light appeared twice in succession. Each light remained on until the correct response was made. Thus, each *S* received two reinforced presentations to two lights and eight to the other two. In the second task each light was presented 10 times, and again, no single light appeared twice in succession. However, only one response was permitted to each light.

In both tasks two of the lights were turned out by one button and two were turned out by

[1] The present paper is a portion of a dissertation submitted to the Graduate College of the State University of Iowa in partial fulfillment of the requirements for the Ph.D. degree in the Iowa Child Welfare Research Station. The writer is indebted to Dr. Alfred Castaneda who directed the dissertation.

[2] Now at Southern Illinois University.

[3] Gratitude is expressed to Mr. David K. Stewart, Elementary School Curriculum Coordinator, Iowa City, Iowa, for his assistance and cooperation in arranging for facilities and subjects.

the other button. Of the four lights used in the first task, one which had appeared eight times and one which appeared twice were selected for re-pairing in the second task.

The instructions for the first task were designed to indicate to *S* that he was to learn which button turned off each light and that if an error was made, correction would be allowed. The second task followed immediately, and neither group was given any information regarding the fact that two of the pairs required re-pairing. The nonpaced group was told to continue but that now correction would not be allowed when an error was made. The paced group was given the same instructions, and in addition, instructions designed to indicate that speed of response was highly important. The *E* demonstrated that the light would be on for 1 sec. and told *S* that a response had to be made before the light went off or it would be counted as an error. However, during the task the interval timer was set to give a 2-sec. exposure of the light rather than the 1-sec. exposure demonstrated. If a response was not made during this interval, the same light was re-presented until a response could be recorded for that trial. Errors due to failure to respond within the interval were not included in the data. No *S* had to have any single light re-presented more than once. Failure to respond within the interval occurred among 14 *S*s for a total of not more than three trials for any one *S*. Of the total number, 9 responses were to the lights requiring no re-pairing, and 13 were to the lights requiring re-pairing.

Results and Discussion

Table 1 presents the mean number of errors made by each group to the four separate stimuli in the first task.

TABLE 1

Errors to the Four Stimuli in the First Task

Level of Habit Strength	Pairing in Second Task							
	Not Re-Paired				Re-Paired			
	Non-paced		Paced		Non-paced		Paced	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Weak	.73	.53	1.12	.59	.73	.71	.80	.75
Strong	.12	.33	.40	.57	.38	.56	.40	.49

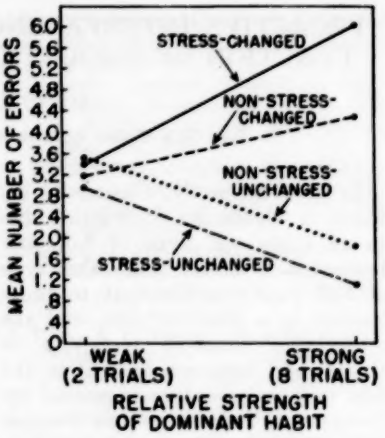


Fig. 1. Mean number of errors for stress and nonstress groups on the stimuli which received different amounts of training and were re-paired or not re-paired.

The group which was paced in the second task made more errors on all four stimuli; however, the differences are significant ($P < .01$) on only those stimuli which were not re-paired in the second task. Since it was predicted that the stress group would make fewer errors than the nonstress group on these two stimuli in the second task, these results would appear to work against the hypothesis.

Figure 1 shows the mean number of errors made in the second task to the various stimuli by the stress and the nonstress groups. It can be noted that, for both groups, more errors occurred on the two stimuli for which the response learned in the first task was no longer correct than to the two stimuli for which the previously learned response was correct. The effectiveness of the amount of training is revealed in the tendency for errors to be greater on the stimulus which had been presented eight times in the first task and for which the previously learned response was no

longer correct, and less on the stimulus which had been presented eight times but for which the previously learned response was still correct. Also, the effects of stress appear to depend on the type of stimulus involved, i.e., the stress group, in comparison to the nonstress group, made more errors on those stimuli for which the previously learned response was no longer correct, but fewer errors on the stimuli for which the same response was correct. Only in the case of the stimuli for which the previously learned response was no longer correct, however, does this latter effect appear to depend on the amount of training.

A summary of an anlysis of variance based on these data, presented in Table 2, indicates that the main effects of stress and amount of training are not significant, while the effects of type of stimulus are highly significant. The interaction of stress and type of stimulus is significant, as well as the interaction of type of stimulus and training level. The triple interaction, indicating that the magnitude of the effects of stress and type of stimulus varied with amount of training, was significant at only the .10 level.

Support for the assumptions that the four stimuli in the second task had associated with them either a correct habit which was relatively weak or strong or an incorrect habit which was relatively weak or strong, is evident in the significant main effects of stimuli and the significant interaction between stimuli and training. Considering the stimuli which required learning of two responses successively, the present findings are in line with those reported by Porter and Duncan (6) and Castaneda (1) that the amount of interference, under such conditions, varies as a positive function of the associative strength of the first learned response. Considering the stimuli for which the same response was correct in both tasks, the present results indicate that the amount of facilitation in the second task varied as a positive function of the associative strength, i.e., the amount of training in the first task. The significant interaction between stress and stimuli is consistent with the two previous studies (2, 3) and in accord with expectations from the assumptions upon which the present study was based. It should be noted here that despite the fact that the stress group made significantly more errors on the unchanged stimuli than the nonstress group in the first task, the effects of stress tended to facilitate performance in the second task on these stimuli.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF ERRORS MADE IN THE SECOND TASK

Source	*df*	*MS*	*F*	*P*
Stress (S)	1	1.44		
Error (b)	49	7.55		
Stimuli (St)	1	176.96	31.10	<.001
Training (T)	1	.08		
St × T	1	165.96	62.39	<.001
St × S	1	31.17	5.48	<.025
T × S	1	5.69	2.14	<.20
St × S × T	1	8.49	3.19	<.10
$Error_1$ (w)	49	5.69		
$Error_2$ (w)	49	2.66		
$Error_3$ (w)	49	2.66		

The expectation that the effects of stress would be a positive function of habit strength, based on the assumption that habit and drive combine multiplicatively, was substantiated in the triple interaction at only the .10 level. On the basis of Fig. 1, the effect appeared more pronounced on the re-paired stimuli than on the stimuli which were not re-paired. It may be noted, however, that a mean of less than 2.0 errors was made by both groups on the one stimulus presented eight times in the first task and not re-paired, hence leaving little room for further improvement.

SUMMARY

The present experiment was concerned with the relation of drive strength to performance

in a complex learning situation. The specific theoretical implication investigated was that the performance of stress (high drive) *Ss* would be poorer than that of nonstress (low drive) *Ss* in a task which involved stimuli for which a response previously learned was no longer correct, but that the performance of the stress *Ss* would be better on the same task involving stimuli for which a response previously learned was still correct. Further, the study investigated the implication that the greater the magnitude of the difference between the strength of the correct and the incorrect habit the greater would be the differences between the performance of the high and the low drive groups.

All 51 *Ss* were given 20 trials on the first task in which four S-R pairings had to be learned. Two of the stimuli were presented more times than the other two. In the second task, 25 *Ss* worked under a stress condition and 26 *Ss* worked under the same conditions as in the first task. One of the stimuli which received a greater amount of practice and one of the stimuli which received a lesser amount of practice were re-paired in the second task, and the other two stimuli were not re-paired.

The results were in agreement with the theoretical expectation that the stress *Ss* would make more errors than the nonstress *Ss* on the re-paired stimuli and fewer errors on the stimuli which were not re-paired. A triple interaction, significant at only the .10 level, was obtained indicating that the magnitude of the effects of stress and type of stimulus vary with the strength of training of the stimulus.

References

1. Castaneda, A. Proactive interference as a function of amount of original training. *Proceed. Iowa Aca. Sci.*, 1955, **62**, 411–414.
2. Castaneda, A. Effects of stress on complex learning and performance. *J. exp. Psychol.*, 1956, **52**, 9–12.
3. Castaneda, A., & Palermo, D. S. Psychomotor performance as a function of amount of training and stress. *J. exp. Psychol.*, 1955, **50**, 175–179.
4. Duncan, C. P. Transfer in motor learning as a function of degree of first-task learning and inter-task similarity. *J. exp. Psychol.*, 1953, **45**, 1–11.
5. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
6. Porter, L. W., & Duncan, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **46**, 61–64.
7. Spence, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In *Kentucky Symposium: Learning Theory, Personality Theory and Clinical Research.* New York: Wiley, 1954.

(Received April 16, 1956)

Journal of Experimental Psychology
Vol. 53, No. 5, 1957

SERIAL EFFECTS IN AUDITORY THRESHOLD JUDGMENTS [1]

WILLIAM J. McGILL

Columbia University

Recent studies have shown that successive threshold judgments in vision are not independent (**2, 10, 11**). The present paper extends this finding to auditory judgments and shows that *S*'s guessing habits are responsible for part of the nonindependence. The data are based on sequences of tones heard against a background of noise. The threshold involved is the threshold for masking. The judgment required was an "absolute" identification of each tone.

Serial correlation in successive "yes-no" judgments suggests autocorrelation in the underlying threshold process. Similar correlations are found in continuous threshold records made with the Békésy audiometer (**1**). Absolute judgments are more complicated because they afford several different opportunities for *S* to be wrong. Nonindependence can creep into such judgments in two ways. First, patterning can appear in a sequence of judgments considered simply as correct or incorrect. Secondly, nonindependence can limit the choice of alternatives among the incorrect judgments. The first type of nonindependence is similar to autocorrelation in continuous threshold recording. The second type of nonindependence appears to depend on *S*'s guessing habits. The present paper is concerned with showing that the nonindependence in auditory absolute judgments is principally of the second type and is, in fact, due to *S*'s guessing habits.

Threshold stimuli are indistinct, and *S* often reports that he is "guessing." The difficulty with such reports is that *S* is seldom a good judge of how well he is doing. A more reliable indication of guessing, if it occurs, can be extracted from the judgments themselves (**9**). For example, one can say that *S* is guessing if his judgments are mutually correlated and independent of stimuli. When this criterion is met the constraints on successive judgments must be internal to *S*. "Guessing" then becomes a label for effects attributable to a broad class of determiners including set, previous training, method of responding, and so on. The essential feature of this class of constraints is that its effects on judgments are the same no matter what stimulus has been presented to *S*. In short, guessing habits act independently of stimuli. It follows that guessing can be measured by measuring the predictability of successive judgments provided that *S* works far enough below threshold to insure that judgments and stimuli are independent. In the present experiment this restriction requires that the level of the noise is high enough to mask the stimulus tones. In the limiting case the tones are omitted altogether.

The present procedure is to vary the intensity of the stimulus tones and observe the predictability of successive judgments, using the guessing series (stimuli omitted) as a baseline. To the extent that the predictability

[1] The research reported in this article was supported jointly by the Army, the Navy, and the Air Force under contract with the Massachusetts Institute of Technology.

of successive judgments, when the tones are more or less audible, resembles the predictability in the guessing series, it may be inferred that *S* is guessing in the threshold series. The results support this inference.

Method

The stimuli were four brief tones at constant intensity but differing in frequency. The frequencies were 890, 925, 970, 1005 cps. A series of these tones was presented one at a time in random order against a noise background. The series was controlled by a tape programmer. Four Hewlett-Packard 200-C audio-oscillators were connected through the programmer to a narrow bandpass filter which went to a mixing circuit where the signals from the oscillators were added to random noise generated by a 6D-4 gas tube. The program sampled a tone from one of the oscillators every 2.8 sec. Thus *S* heard one of four randomly chosen tones every 2.8 sec. The tone lasted for .5 sec. and was superimposed on a steady background of noise. While the tone sounded, a light flashed directly in front of *S* to signal the presentation of the stimulus. Each oscillator was set to generate one of the four tones. The larger gap between the middle tones helped to equalize the difficulty of the judgments. Listening was binaural. The earphones were type PDR-8. Measurements of signal and noise intensity were made at the input to the earphones. Stimulus intensity is reported as the ratio of signal power to the noise power per cycle or spectrum level of the noise. The intensity of the tones was fixed, and signal-to-noise ratio was changed by varying the noise.

The *S* was given preliminary training in which he learned to make "absolute" judgments (identifications) of the four tones by pressing four response keys. He was instructed to guess when in doubt. During training *S* was corrected freely, but no information about the accuracy of his judgments was given during experimental sessions. Before each session *S* was required to pass a pretest consisting of 100 tones. The passing criterion was 90% correct. In the pretest no masking noise was added. However, the narrow gap between adjacent frequencies produced confusions even in the absence of masking noise. Frequent retraining was necessary in order to maintain the pretest standard of accuracy.

Each *S* was tested at 15, 22, 25, 30, 35 db. Each test consisted of a series of 250 stimuli. In addition, a guessing series was obtained on each *S*. This series resembled the others except that the stimuli were disconnected from the input to the earphones. The tones appeared to be masked completely by noise, and *S* indicated his guess by pressing one of the response keys whenever the stimulus light flashed. The *S* was not informed that there were no stimuli.

Tests were run in a modified counterbalanced order. The series at 15 db. and the guessing series were always done last. The other four intensities were arranged in a latin square with *S*s.

Four M.I.T. undergraduates served as *S*s. Each *S* provided 1500 judgments.

Results

Figure 1 shows the combined psychometric function for intensity of the tones above the masking noise. The curve in Fig. 1 is the average of the curves determined at 890, 925, 970, and 1005 cps. Each point is based on 250 judgments by each of four *S*s. The function in Fig. 1 is pushed down or compressed from above by the difficulty level of the absolute judgments, since the pretest criterion was barely met in most sessions. The curve is also compressed from below by the probability of chance success—which in this case is .25. These compressions make the choice of a unique threshold point somewhat arbitrary. The usual pro-

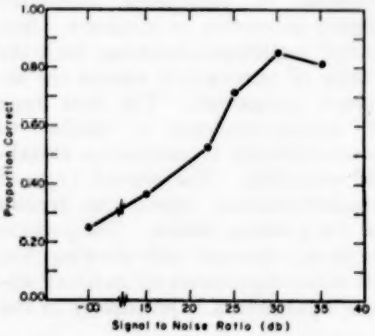


Fig. 1. Psychometric function for absolute judgments of brief tones superimposed on background of masking noise. Data are averaged over four tones: 890, 925, 970, 1005 cps.

cedure is to adjust the observed proportions to allow for chance success. The proportion corresponding to threshold is then computed from:

$$\frac{k+1}{2k} \qquad k \geq 2,$$

where k is the number of forced choice alternatives. In Fig. 1 this procedure fixes the threshold at .63, or approximately 24 db. which is considerably higher than the figure of 18 db. found by the method of adjustment (6). This seems to reflect the difficulty of identifying four tones instead of one.

The chance correction, however, is highly questionable with the present data because it assumes that successive judgments are independent, and it will be demonstrated that this is not the case. In fact, the threshold level is so high as to suggest that some odd things may occur in absolute judgments. The S is forced to do a lot of guessing, and this may somehow affect his sensitivity. In any case, the psychometric function is quite fuzzy because successive judgments are not independent. A better picture of what is going on is achieved by introducing the idea of response predictability. Wherever the four tones are correlated with the responses made by S, the stimuli may be said to predict the responses up to the extent of their influence. Similar predictions are possible from S's previous responses. The total predictability of S's current judgment may be expected to involve elements of both classes of predictors.

The principal results trace shifts in predictability from the stimulus to S's last previous response (hereafter called the "presponse") as the level of the masking noise and, consequently, the difficulty of the judgment varies in the region of the threshold.

In order to measure the effects of two predictors simultaneously, the data are treated by uncertainty analysis as described by McGill (8) and Garner and McGill (5). This method was also used by Garner to study presponse effects in loudness judgments (4). In the present paper the terms "information" and "uncertainty" are used interchangeably. These two terms sometimes describe different aspects of information measures (5), but the differences are unimportant here.

In the present application of uncertainty analysis, the data are arranged in a three-dimensional contingency table. Each response is categorized according to S's identification of it, the stimulus actually presented, and S's presponse. Let w_i represent the ith category of presponses; x_j, the jth category of stimuli; and y_k the kth category of S's identifications. Then n_{ijk} is the number of responses in the cell of the contingency table determined by the intersection of categories, i, j, and k. Marginal sums are designated by replacing indexes with dots. For example, $n_{.jk}$ is the number of times S puts stimulus j into response category k. The total uncertainty (information) in S's responses is measured as follows:

$$U(y) = -\sum_{k}\left[\frac{n_{..k}}{n}\right]\log_2\left[\frac{n_{..k}}{n}\right]. \quad (1)$$

In Equation 1, n is the total number of responses. It is easily shown that

$$U(y) = U_x(y) + U(y{:}x), \quad (2)$$

where $U_x(y)$ is the uncertainty remaining in S's responses when the stimulus is known, and $U(y{:}x)$ is the uncertainty in S's responses that is predicted by the stimulus. $U(y{:}x)$ is called transmitted information or contingent uncertainty (5) and is similar to the variance of points on a regression line in metric data. The uncertainty contingent on stimuli, i.e., the last term in Equation 2, is computed as shown in Equation 3:

$$U(y{:}x) = -\sum_{j,k}\left[\frac{n_{.jk}}{n}\right] \times \log_2\left[\frac{n_{.j.}\, n_{..k}}{n\, n_{.jk}}\right]. \quad (3)$$

The last step in the analysis is to extract the presponse component from

the error term in Equation 2. This is done as follows:

$$U_s(y) = U_{ws}(y) + U_s(y:w). \quad (4)$$

The new error term is $U_{ws}(y)$. The response component is the second term on the right of Equation 4. The formula is given in Equation 5:

$$U_s(y:w) = -\sum_{i,j,k}\left[\frac{n_{ijk}}{n}\right] \times \log_2\left[\frac{n_{ij.}\, n_{.jk}}{n_{ijk}\, n_{.j.}}\right]. \quad (5)$$

The quantity in Equation 5 is the uncertainty contingent on previous responses. It is determined separately for each stimulus category and then averaged. The averaging essentially partials out the stimulus variable and makes Equation 5 a close relative of partial correlation.

The response uncertainty predicted by the four stimulus tones (Equation 3) is presented in the column headed "Stimulus Component" in Table 1, which shows how the predictability of *S*'s judgments changes with signal-to-noise ratio. All data are measured in bits and averaged over the four *S*s.

Response uncertainty (i.e., total uncertainty determined from Equation 1) is stable at approximately two bits at each stimulus intensity and in the guessing series. This means that

TABLE 1

CHANGES IN STIMULUS COMPONENT OF UNCERTAINTY IN *S*'S JUDGMENTS PRODUCED BY CHANGING INTENSITY OF STIMULUS TONES

Relative Intensity (db.)	Stimulus Component	Total Uncertainty
Guessing series	.03	1.97
15	.13	1.97
22	.58	2.01
25	1.05	1.97
30	1.37	1.99
35	1.30	2.01

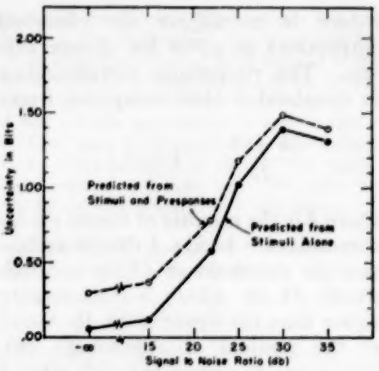


FIG. 2. Uncertainty in *S*'s judgments predicted jointly from stimulus tones and *S*'s last previous response.

*S*s used the four frequency categories equally often, whether they heard the tones or not. Two measures of total uncertainty in Table 1 exceed two bits, indicating an occasional failure to respond. Failure to respond was treated as a fifth response category.

As suggested previously, *S*'s judgments are predictable from stimuli and presponses jointly. Curves representing the joint prediction and the prediction based on stimuli alone are compared in Fig. 2. The joint prediction in Fig. 2 is the nonmetric analogue of a multiple correlation derived from the three-dimensional data matrix. The uncertainty component predicted from stimuli is the one just presented in Table 1. It is based on the two-dimensional matrix relating stimuli and *S*'s responses. This component increases with intensity in about the same way as the proportion of stimuli correctly identified. Contingent uncertainty, however, is not compressed from below by chance successes. When *S* guesses, the stimulus component drops to zero because stimuli and responses are independent. Actually this is not strictly true, since it turns out that

contingent uncertainty is biased away from zero even when stimuli and responses are independent. The bias, however, has been shown (7) to be approximately equal to

$$\frac{(j-1)(k-1)}{1.3863n},$$

where j is the number of stimulus categories, k is the number of response categories, and n is the number of responses. In the present data the bias amounts to .026 bits. This is roughly the magnitude of the stimulus component shown for the guessing series in Table 1. In the guessing series a stimulus tape was placed in the programmer each time, and *S* responded in the usual way. However, the programmer was disconnected from *S*'s earphone. Consequently, any information predicted from stimuli is due either to good luck or to ESP, neither of which seems to have occurred.

The component of response uncertainty predicted from presponses is measured with stimuli held constant. As previously noted, this uncertainty component is determined separately for each stimulus and then averaged.

TABLE 2

CHANGES IN PRESPONSE COMPONENT OF UNCERTAINTY IN *S*'S JUDGMENTS PRODUCED BY CHANGING INTENSITY OF STIMULUS TONES

Relative Intensity (db.)	Size of Presponse Component		
	Obtained	Maximum	% Maximum
Guessing series	.29	1.94	15
15	.22	1.84	12
22	.20	1.43	14
25	.13	.92	14
30	.12	.62	19
35	.11	.71	15

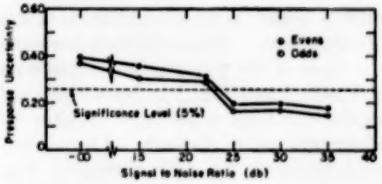


FIG. 3. Uncertainty in *S*'s judgments predicted from *S*'s last previous response, with stimulus held constant. Presponse influences on judgments are presented separately for odd- and even-numbered responses.

Presponse components obtained at each stimulus intensity are shown in Table 2. The presponse effect is largest in the guessing series and declines slowly up to 22 db. There is a sharper drop between 22 and 25 db., after which the presponse effect is stable.

Figure 3 shows the presponse component at each level of stimulus intensity. Separate curves are presented for the odd-numbered and even-numbered responses. This procedure is followed to prevent counting the same response twice—once as a presponse and again as a response. In Fig. 3 the odd-numbered responses are presponses for the function based on even-numbered responses and vice versa. Significance level is obtained from the likelihood ratio test for association between presponses and responses with stimuli held constant. The critical value of chi square at the 5% level is used to calculate a presponse component significantly different from zero, and a dashed line is drawn across Fig. 3 at this point. The presponse component of response uncertainty drops below the critical value at 25 db. and remains below thereafter. Additional calculations recommended by Miller and Madow (7) reveal that the observed presponse component above 25 db. is about equivalent to the statistical bias ex-

pected when successive responses are independent. The difference between the sizes of the presponse components in Table 2 and Fig. 3 is due to inflated statistical bias when the number of observations is halved. The difference between the two curves in Fig. 3, however, is not easy to explain. It may reflect a periodicity (of the autocorrelational type) in the data.

Table 2 also presents the maximum size and the relative size of the presponse component. The presponse component is extracted from the response uncertainty when the stimulus is known, i.e., from *S*'s identification errors. Equation 4 shows that the latter determine the maximum value of the presponse component for any given stimulus intensity. The maximum, $U_x(y)$, is computed by subtracting the stimulus component from the total uncertainty in Table 1. This difference is given in Table 2. The relative size of the presponse component is determined from the ratio of the obtained size to the maximum size. This ratio is practically constant. There is nothing in the analysis that forces this result. The picture appears to be the following. In the two-dimensional matrix relating stimuli and responses, the number of *S*'s responses that fall in the correct-response cell increases with S/N ratio. But the distribution of *S*'s *incorrect* responses remains stable, so that a constant proportion of these incorrect responses can be predicted from presponses. Hence, the patterning of incorrect responses seems to be determined by presponses and reflects *S*'s nonrandom guessing habits. The type of patterning is not revealed by the analysis, but the presponse-response matrices disclose that *S* tends to avoid repeating his last response.

Discussion

This experiment demonstrates that presponses have a marked effect on absolute judgments when stimulus intensity is at or slightly below threshold. Hence, successive judgments are not independent. Furthermore, the effect blends smoothly with *S*'s judgments in the guessing series. The combined effects appear to reflect a lawful process in absolute judgments. Consequently, it is worth while to try to identify the variables that affect *S*'s guessing habits.

One variable appears to be the number of response alternatives, especially when *S* has been set by training or instructions to respond randomly. Experiments with other guessing series, not reported here, indicate that presponse influences are small with two-category forced choice judgments. Garner (4) has shown that presponse influences change systematically with the number of alternatives between 5 and 20 categories. A second variable is the time interval between successive judgments. In a check experiment at 25 db. on the *S*s in this experiment, it was observed that the presponse component declined by small amounts but consistently as the interstimulus interval increased. Tests were made at intervals of 2.8, 5.6, 8, and 11 sec. Day's results (3) show that the most important presponse influences are found with time intervals shorter than 2.1 sec.

A third variable controlling presponse influences is the difficulty level of the absolute judgments. The variation of the presponse effect in the vicinity of the threshold demonstrated in this experiment illustrates this control. In another check experiment on *S*s in this experiment, the difficulty level of the judgment was eased by increasing the highest frequency from 1005 to 1075 cps. Tests were run at 20, 25, 30, and 35 db. Information transmitted from stimuli exceeded the value reported in Table 1 at each intensity. Presponse components were correspondingly lower at each intensity.

In this study presponse influences were limited to a single prior response. Influences from earlier responses were thus relegated to the error term of the information analysis. It is possible that genuine long-term influences exist and that their extent depends on the variables described above. Long-term presponse influences have been claimed, but an important methodological consideration has been overlooked. Suppose only the last previous response affects a judgment. This influence will be propagated spuriously over the entire series of judgments because, if response *a* affects response *b*, and *b* affects *c*, then *a* will also appear to affect *c* unless *b* is partialed out. Consequently autocorrelational studies of response sequences reveal little about the structure of long-term influences. In fact, autocorrelations may be very misleading since they often indicate much more complicated structures than actually exist. What is required is a study of long-term influences with intervening effects partialed out, but data of this kind are easier to conjecture than to obtain.

Summary

Four *S*s learned to identify tones at 890, 925, 970, 1005 cps. The *S*s then judged sequences of 250 brief stimuli delivered in random order at constant intensity. The difficulty of the judgment was varied by varying the level of a background masking noise. The data were then analyzed to determine the joint effect of the stimulus and *S*'s guessing habits (called presponse effects) in determining judgments. Results were as follows:

1. A marked presponse effect was observed for signal-to-noise ratios lower than 22 db. This corresponds to judgments at or below the average value of the threshold.

2. The presponse influence was greatest when no stimuli were presented. (The *S*s were unaware that there were no stimuli.) The presponse influence declined gradually as tones were introduced and made more intense.

3. Information analysis showed that the frequency of *S*'s judgment errors was controlled by stimulus intensity, while the pattern of errors was governed by *S*'s guessing habits.

References

1. Békésy, G. von. A new audiometer. *Acta Oto-Laryng., Stockh.*, 1947b, **35**, 411–422.
2. Collier, G. Probability of response and intertrial association as functions of monocular and binocular stimulation. *J. exp. Psychol.*, 1954, **47**, 75–83.
3. Day, W. Serial nonrandomness in auditory thresholds as a function of interstimulus interval. *Amer. J. Psychol.*, 1956, **69**, 387–394.
4. Garner, W. R. An informational analysis of absolute judgments of loudness. *J. exp. Psychol.*, 1953, **46**, 373–380.
5. Garner, W. R., & McGill, W. J. The relation between information and variance analyses. *Psychometrika*, 1956, **21**, 219–228.
6. Hawkins, J. E., & Stevens, S. S. The masking of pure tones and speech by white noise. *J. acoust. Soc. Amer.*, 1950, **22**, 6–13.
7. Miller, G. A., & Madow, W. *On the maximum likelihood estimate of the Shannon-Wiener measure of information.* Cambridge: USAF Cambridge Res. Center, Tech. Rep. No. TR-54-75, 1954.
8. McGill, W. J. Multivariate information transmission. *Psychometrika*, 1954, **19**, 97–116.
9. Senders, V. L., & Sowards, A. Analysis of response sequences in the setting of a psychological experiment. *Amer. J. Psychol.*, 1952, **65**, 358–374.
10. Verplanck, W. S., Collier, G. H., & Cotton, J. W. Nonindependence of successive responses in measurements of the visual threshold. *J. exp. Psychol.*, 1952, **44**, 273–282.
11. Verplanck, W. S., Cotton, J. W., & Collier, G. H. Previous training as a determinant of response dependency at the threshold. *J. exp. Psychol.*, 1953, **46**, 10–14.

(Received March 21, 1956)

Journal of Experimental Psychology
Vol. 53, No. 5, 1957

THE ABSTRACTION AND CONCEPTUALIZATION OF FORM, COLOR, AND NUMBER [1]

JOACHIM F. WOHLWILL

University of California

The question of the relative facility with which concepts of form, color, and number are attained in different situations has been extensively investigated by Heidbreder and her co-workers (**8, 10, 11**), and subsequently by Grant and his associates (**3, 4**) using the Wisconsin Card Sorting Test (WCST). The results from these studies have shown, in general, that abstract forms are more readily conceptualized than colors or numbers, although with respect to the latter, the two series of studies have yielded somewhat contradictory results.

Heidbreder (**6**) has maintained that such studies give evidence of a dominance-hierarchy of concepts, extending from concepts of concrete objects to those of number, according to their degree of "thing-character." The nature and origin of this hierarchy remains, however, uncertain in a number of respects. One of these concerns the place of perception in this hierarchy of concepts. Specifically, is readiness for conceptualization primarily a function of facility of perceptual abstraction of a given stimulus attribute, or are conceptual predispositions responsible for it which are further removed from the perceptual process?

While Heidbreder's experiments (**8**) actually show a *decreasing* influence of the hierarchy of concepts as the role of direct perception in a given task increases and therefore argue against the former of the above alternatives, she nevertheless goes to some length in one of her more recent articles to explain that the order of dominance of concepts as advocated by her is primarily one of perceptual rather than conceptual processes (**9**, pp. 117–118).

One way of exploring this problem further would be to compare the order of dominance of stimulus attributes in two contrasting tasks involving identical stimulus materials—one of a problem-solving nature, in which hypotheses based in part on *S*'s predispositions might be displayed, the other of a perceptual nature, in which the response to be made is completely defined for *S*.

A concept formation experiment closely similar to the WCST, but using probabilistic reinforcement of the different concepts, was selected for the first of these tasks. By using probabilistic reinforcement and thus preventing *S* from finding a consistently correct mode of response, it was hoped to maximize any tendencies on the part of *S* to respond to one rather than another of the various stimulus-cues.

The second task, a card-sorting task in which speed of response to a predetermined stimulus dimension of the card was measured, provided for a comparison of these same cues in terms of their relative ease of abstraction. As used in this paper, this term *abstraction* is taken to refer to that aspect of the discrimination process which involves the response to a given sensory dimension in the

[1] The writer gratefully acknowledges the generous advice and criticism offered by Dr. Donald A. Riley.

presence of variation of the stimuli in other dimensions. An operational definition for the measurement of speed of abstraction in this sense is presented later.

Method

Part I

This part of the experiment concerns the comparison of concepts of form, color, and number in a probabilistic concept formation task.

Stimuli.—Forty-eight response cards from the WCST were selected, together with the four WCST stimulus cards. These varied in form (triangle: T; star: S; cross: X; circle: C), color (red: r; blue: b; green: g; yellow: y) and number (one, two, three, four). Every response card matched any given stimulus card on no more than one dimension.

Procedure.—The *S* was seated in front of a sorting board with four compartments, behind which the four stimulus cards had been placed in the following order: 2gS, 4bC, 3yX, 1rT. He was given the deck of 48 response cards, with the following instructions:

" . . . You are to sort this deck of cards in front of you one by one into the compartment underneath the stimulus card with which you think the card to be sorted belongs. After you have sorted each card, I will tell you whether you were right or wrong. Your purpose should be to get as many correct as you can. While this is not a timed task, you should not spend too much time on any one card, but rather try to sort at a fairly even rate. . . . Always wait for my announcement of 'right' or 'wrong' before proceeding to sort the next card."

Conditions of reinforcement.—Reinforcement schedules were prepared in advance, such that for two-thirds of the cards in the deck one of the three stimulus dimensions would be the correct basis for sorting, while reinforcements on the remaining third of the deck were divided up equally between the other two dimensions. For one-third of the *S*s the "primary" concept (the one most often reinforced) was form, for one-third it was color, and for one-third number. As an example of this arrangement, with number as the primary concept, the correct basis for sorting on the first 10 cards varied among Form (F), Color (C) and Number (N) as follows: N, N, N, F, N, N, C, N, F, N.

Part II

In Part II the speed of abstraction of the form, color, and number aspects of the WCST cards was compared. It could not be assumed, however, that the particular forms represented in these cards were equal in discriminability to the particular colors or numbers represented, i.e., that the perceptual information transmitted by each stimulus-cue was the same. In order to control for this factor, these forms, colors, and numbers had to be presented independently of variations in the other dimensions, the resultant sorting times serving as a base line against which to evaluate differential speed of *abstraction*.[2] This was the purpose of the "simple" cards, presently to be described.

Stimuli.—The cards used in this part of the experiment were of three kinds: "simple," "mixed," and "practice."

"Simple" cards: Three sets of 24 simple cards were used, as follows: a set of 24 simple-form cards, varying only in form and consisting of six replicas of each of the four geometric shapes of the WCST cards, centered on the card in solid black; a set of 24 simple-color cards, varying only in color and consisting of six replicas of each of the four colors of the WCST, in the form of squares centered on the card; lastly, a set of 24 simple-number cards, varying only in number and consisting of six instances of from one to four black squares in different configurational patterns.

"Mixed" cards: Six sets of 24 mixed cards were used for the determination of differential speed of abstraction, two sets to be sorted by form, two by color, and two by number, these sets being made up of cards from the WCST.[3]

"Practice" cards: Two sets of 24 practice cards were used, these being identical with one of the four WCST stimulus cards in every respect, so that sorting them did not involve abstraction of any one stimulus dimension.

Each of the above sets of response cards had associated with it a set of four stimulus cards. For the simple card sets these contained the four types of symbols represented in the respective

[2] The validity of sorting times as a measure of the perceptual information transmitted by a set of stimuli has been demonstrated, at least for the continuum of number, by Crossman (2).

[3] These mixed cards were actually of two different kinds, representing the "single cue" and the "double cue" condition, respectively. A card in the single cue sets agreed with the stimulus card with which it was to be sorted only on the dimension indicated as the basis for sorting, while a card in the double cue sets agreed with its stimulus card on that dimension and on an additional one which changed from card to card. But since the results with respect to this variable were inconclusive, further mention of it has been omitted from this paper.

set (e.g., a black triangle, circle, cross, and star, for the simple-form set); for the mixed sets and the practice sets the stimulus cards of the WCST (cf. Part I) were used.

All cards were of thin cardboard, 3 × 3 in.

Procedure.—Part II followed directly upon completion of Part I. The *Ss* were divided into three groups, according to order of sorting on the three dimensions (FCN, CNF, and NFC).

All *Ss* started with one of the practice sets, followed by a series of six decks of cards in different orders. For Group FCN the order was: simple-form; mixed, to be sorted by form; simple-color; mixed, to be sorted by color; simple-number; mixed, to be sorted by number. The order was changed accordingly for *Ss* in the other groups, simple-card sorting for each dimension always preceding mixed-card sorting. Lastly, all *Ss* received the second of the practice sets.

The *S* was asked to sort each deck according to the dimension indicated by *E*, and to do so as fast as possible, consistent with accuracy. Sorting times were recorded to the nearest .1 sec. by means of an electric timer.

Subjects.—The *Ss* were 24 male and 24 female undergraduates, enrolled in psychology courses at the University of California. Due to a slight revision in experimental procedure for Part II, data for that part are based on only 18 *Ss* of each sex.

Results

Part I

The main measure in the analysis of the results for Part I is the number of cards sorted according to the primary concept, i.e., the one reinforced on two-thirds of the cards. The means of the form, color, and number groups for successive blocks of 12 trials are presented in Table 1. As is shown in this table, *Ss* responded more consistently to the primary cue when it was form or number than when it was color, particularly in the first stages of the task.

TABLE 1

Mean Number of Cards Sorted by Primary Cue

Primary Cue	Block of Trials				All Trials
	1	2	3	4	
Form	8.9	9.3	8.5	9.3	36.0
Color	5.4	6.8	7.4	8.1	27.7
Number	7.1	8.9	8.6	9.2	33.8

The statistical analysis of these data presents certain difficulties, due to the uncertain form of the distribution of the scores. These are characterized both by marked heterogeneity of variance and fairly radical departure from normality. It was therefore decided to resort to a nonparametric analysis of variance recently proposed by Wilson (**14**), based on the number of cases above and below the median in each cell. Due to the lack of independence of the scores for successive blocks of trials, this analysis had to be limited to the total number of cards out of 48 sorted by the primary cue. The resultant χ^2 for cues was 9.50, significant beyond the .01 level.[4]

In order to corroborate these findings, the same analysis of variance was carried out on the maximum number of *successive* cards sorted by the primary cue for each *S*. The median length of such runs for each cue was: Form, 13; Color, 9.5; Number, 14. The corresponding χ^2 for cues is 12.27, significant beyond the .01 level.

On both measures, then, color is less consistently responded to than either form or number when these are the primary reinforced cues, while form and number appeared to be about equally effective.[5]

[4] Although it is felt that the nature of these scores demands the nonparametric analysis, mention should be made of the results obtained from a Fisherian analysis of variance on these same scores. This yielded an *F* for sorting categories of 2.91, with 3.22 being required for the .05 level of significance. A subsequent *t* test for the difference between form and color resulted in a *t* of 2.33, significant between the .05 and .02 levels.

[5] A breakdown of the data by sex discloses that for the male *Ss* the response to number actually exceeded that to form in consistency in

It may perhaps aid in the understanding of these results to give a brief qualitative description of the mode of performance most commonly observed in this situation. After the first two or three nonreinforced cards, *S* would usually try out a number of different hypotheses in the search of one which might pay off more consistently. These consisted not only of the three main cues, but also of combinations of these, as well as of positional and configurational cues. As the sorting went on, however, the primary cue was usually responded to more consistently regardless of nonreinforcements, this being especially true for form and number (thus the rather high averages on the maximum-run measure). As a whole, this type of response bears rather little resemblance to the probability-matching typically observed in two-choice experiments with probabilistic reinforcement. The difference is due largely, one suspects, to the much larger variety of possible hypotheses in this situation and the opportunity given to *S* to test them, by letting him sort at his own rate.

Part II

The results from Part II are summarized in Table 2. With respect to the simple-card sorting times, the significant differences among cues presumably reflect differences in discriminability of the particular instances of each category represented in the WCST cards. These differences clearly contaminate the mixed-card sorting times, which are thus of little value in themselves. Attention will be focused, therefore, on the "abstraction scores." These were obtained for each *S* by subtracting his simple-card times from his mixed-card times. The resulting difference scores indicate the extent to which the irrelevant dimensions interfered with sorting on a given category, i.e., the relative difficulty of the abstraction of these stimulus attributes. They cannot, however, be taken in any absolute sense (i.e., where negative, as indicating facilitation from the other dimensions), due to probable time-order effects of proactive facilitation from the simple cards to the mixed cards.

TABLE 2

MEAN SORTING TIMES (IN SEC.) AND ABSTRACTION SCORES, BY SORTING CATEGORY

Sorting Category	Simple Cards	Mixed Cards	Abstraction Scores
Form	22.3	24.9	2.6
Color	20.4	19.8	− 0.6
Number	24.5	24.5	− 0.0
F for cues	44.3*	56.8*	16.1*

* $P < .001$.

As seen in Table 2, these abstraction scores show unequivocally that color is more easily abstracted than form, in contrast to the findings from Part I. Number occupies here an intermediate place, although much closer to color than to form.

DISCUSSION

It is apparent that there is little if any correspondence between the results of

the last half of the task, with a mean of 21 out of the last 24 cards being sorted by number, compared to 19.5 sorted by form. This is of interest in the light of Grant and Curran's (3) finding that number sorting gives rise to a disproportionate amount of errors of perseverance on the WCST, after the correct sorting category is shifted. However, the female *S*s in the present sample failed to show any similar adherence to number as a primary cue. Since the data, moreover, do not permit a trend analysis, the results can only be held to be suggestive as regards this possible characteristic of number.

With respect to the number cue, there is actually some reason to doubt that the abstraction scores represent an altogether valid measure of relative ease of abstraction. It seems likely that the rather high simple-number times were due to some extent to the 2–4–3–1 order in which the stimulus cards were arranged (this being necessary in order to prevent *S*s from ignoring the stimulus cards when sorting for number, as they might do if these were arranged in numerical order). Although this factor was still operating in the mixed-card sorting, the practice on the preceding simple-number cards probably served to weaken it considerably, which may help account for the mean abstraction score of zero for number.

the two parts of this experiment, particularly as regards the relative position of color and form. In the first part, form is "dominant" over color, in the sense of representing a more stable and consistent conceptual category. In the second part, on the contrary, color is considerably easier to abstract.

This latter finding is, incidentally, in line with that obtained in a previous study by Hamlin, Stone, and Moskowitz (5), although in this study differences in discriminability of the form- and color-instances used were not controlled. Ths most plausible explanation for thie superiority of color probably lies in perceptual factors inherent in these stimulus dimensions, such as the redundancy in the stimulus with respect to the relevant dimension, this being clearly greater for color than for form.

Yet the fact remains that in a problem-solving situation such as Part I of the experiment, or a task like the WCST, as well as in Heidbreder's "modified memory" series (10) and other concept-learning experiments (e.g. 12), form is a preferred aspect of the stimulus to which to respond. Although the explanation for this dominance of form is far from clear, it seems reasonable to attribute it to the operation of response-sets which the individual brings to the situation and which have been formed during the course of his mental development and experience. Some support for such a view derives from genetic studies of color-form preference, which seem to point to a change from color- to form-dominance in early and later childhood (cf. 1, 13).

The relevance of the present results to Heidbreder's theoretical formulation of the conceptual process deserves discussion. One might take them to support Heidbreder's distinction between situational and organismic determinants of conceptual performance, although the operational basis for such a distinction has yet to be established. The important point to note, in any event, is that if one may speak of an organismic order of dominance of concepts at all, it must be considered one of conceptual rather than perceptual processes.

Perhaps the most questionable aspect of Heidbreder's conception, however, is her notion of the "thing-character" of a concept as the basis for this order of dominance. This is most apparent in regard to the place of number, a conceptual category relegated to the bottom of the hierarchy by Heidbreder, due to the inherently symbolic character she attributes to it (7, pp. 218–220). Actually, the results of Part I show the number category to be closely comparable to form, especially after the initial stage of the task, as witness the medians on the maximum-run measure.

In defense of Heidbreder it may be argued that the WCST cards do not contain instances of concrete objects, with respect to which her thing-character hypothesis was most consistently confirmed in her own experiments. But the difficulties encountered in explaining the unstable place of color in her results (cf. 11, pp. 65–66), as well as the above-mentioned conflicting evidence with respect to number, would seem to cast doubt on the adequacy of "thing-character" as an explanatory concept of more general applicability.

In conclusion, the marked discrepancy between the results of the two parts of this experiment suggests the advisability of distinguishing more clearly than has usually been done heretofore between two evidently rather different cognitive processes: one, *abstraction*, being a selective response to a given aspect of a stimulus which is probably involved to a greater or lesser extent in any act of perceptual discrimination; the other, *conceptualization*, being perhaps more profitably considered as a process of mediated generalization, and as such being primarily a function of the prior experience and stage of mental development of the individual.

Summary

The relative strength of response to the form, color, and number dimensions of the Wisconsin Card Sorting Test cards was compared in two

tasks: a problem-solving type task using probabilistic reinforcement of these cues, and a perceptual abstraction task involving speed of sorting on the three stimulus dimensions.

The results indicated that when form or number were the primary reinforced concepts, they were more consistently responded to than when it was color. On the other hand, form was much less readily abstracted than color or number in the second task, this being true even after differences in discriminability of the particular colors, forms, and numbers represented in these cards had been taken into account.

These contrasting results suggest the need for differentiating between a perceptually determined process of abstraction and a supra-perceptual process of conceptualization. With respect to the latter, certain difficulties for Heidbreder's formulation of a dominance-hierarchy of concepts based on "thing-character" are pointed out, the superior showing of number being especially noteworthy.

References

1. Brian, C. R., & Goodenough, F. L. The relative potency of color and form perception at various ages. *J. exp. Psychol.*, 1929, **12**, 197–213.
2. Crossman, E. R. F. W. The measurement of discriminability. *Quart. J. exp. Psychol.*, 1955, **7**, 176–196.
3. Grant, D. A., & Curran, J. F. Relative difficulty of number, form and color concepts of a Weigl-type problem using unsystematic number cards. *J. exp. Psychol.*, 1953, **43**, 408–412.
4. Grant, D. A., Jones, O. R., & Tallantis, B. The relative difficulty of number, form and color concepts of a Weigl-type problem. *J. exp. Psychol.*, 1949, **39**, 552–557.
5. Hamlin, R. M., Stone, J. T., & Moskowitz, M. J. Rorschach color theories as reflected in simple card sorting problems. *J. proj. Tech.*, 1955, **19**, 410–415.
6. Heidbreder, E. Toward a dynamic psychology of cognition. *Psychol. Rev.*, 1945, **52**, 1–22.
7. Heidbreder, E. The attainment of concepts: II. The problem. *J. gen. Psychol.*, 1946, **35**, 191–223.
8. Heidbreder, E. The attainment of concepts: VII. Conceptual achievements during card-sorting. *J. Psychol.*, 1949, **27**, 3–39.
9. Heidbreder, E. Experiments by Dattman and Israel on the attainment of concepts. *J. Psychol.*, 1952, **34**, 115–136.
10. Heidbreder, E., Bensley, M. L., & Ivy, M. The attainment of concepts: IV. Regularities and levels. *J. Psychol.*, 1948, **25**, 299–329.
11. Heidbreder, E., & Overstreet, P. The attainment of concepts: V. Critical features and contexts. *J. Psychol.*, 1948, **26**, 45–69.
12. Kendler, H. H., & D'Amato, M. F. A comparison of reversal and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49**, 165–174.
13. Lindberg, B. J. *Experimental studies of colour and non-colour attitude in school children and adults.* Copenhagen: Levin & Munksgaard, 1938.
14. Wilson, K. V. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, **53**, 96–101.

(Received April 26, 1956)

Journal of Experimental Psychology
Vol. 53, No. 5, 1957

THE EFFECT OF DRIVE REVERSAL ON LATENCY, AMPLITUDE, AND ACTIVITY LEVEL [1]

ROBERT H. DAVIS[2]

Michigan State University

In a study recently reported by Deese and Carpenter (**4**), it was incidentally observed that shifting animals from a low drive to a high drive produced a sudden drop in running time, but that shifting animals from a high drive to a low drive failed to produce a significant increase in running time. This unilateral effect due to shift in drive level is called here the drive-reversal effect. The present study is concerned with the drive-reversal effect for thirst as well as hunger and with respect to three response measures, namely, latency, amplitude, and general activity level.[3]

METHOD

Subjects.—The *S*s were 36 male albino rats from the colony maintained by the Department of Psychology of Michigan State University. Age ranged from 85 to 125 days, with a mean of approximately 90 days.

Apparatus.—The apparatus permitted the joint measurement of three aspects of response: (*a*) latency, (*b*) amplitude or force, and (*c*) prior activity level. The measures of activity level and force differed from previous ones in that the force of a skeletal rather than an autonomic response was measured, and in that activity level was measured in the same apparatus as the other responses were.

Figure 1 presents a cross section of the apparatus. At the bottom of the activity chamber there was a false floor supported by three springs and a rubber ball at its exact center. Small rubber balls were attached to the corners of this false floor and prevented the floor from tipping more than $\frac{1}{4}$ in. The interior of the chamber was painted flat gray and was illuminated by a 7-w. bulb located at one end of the box and covered with opal flashed glass. A guillotine door at the other end of the activity chamber, when raised, gave access to a hinged 4-in. × 2-in. push-panel and activated a Standard Electric timer.

Activity level was measured by a 115-v. impulse counter connected in series to a GE 2-36KRI mercury switch which was suspended beneath the floor in such a manner that movement by *S* caused the mercury in the tube to move, momentarily making and breaking the circuit in accordance with the strength and the number of the movements. The initial force of the push-panel was determined in the following manner. A thin, metal rod, hinged at the top

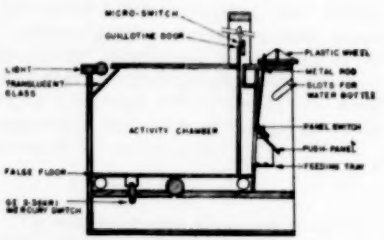


FIG. 1. Cross-sectional drawing of experimental apparatus.

[1] This paper is based upon a portion of a dissertation submitted to the Department of Psychology of Michigan State University in partial fulfillment of the requirements for the Ph.D. degree. The author is indebted to Dr. M. Ray Denny for advice and assistance throughout the investigation.

[2] Now at The RAND Corp., Santa Monica, California.

[3] The original investigation (**3**) was primarily directed at answering a number of methodological questions in research on motivation. The main problem was to determine an amount of water reward which was equivalent to a given amount of food reward as an incentive. Incentive value was assessed in terms of general activity level, and the latency, amplitude, and resistance to extinction of an instrumental response. For all of these measures, it was found that .08 gm. of food was approximately equivalent to .20 cc. of water for both high and low drive. With respect to latency and activity level this is indicated in Tables 1 and 2. Reliability estimates in the form of correlation coefficients for activity and force of response were as follows: activity .62, and force .86.

of the push-panel was twisted so as to extend to the back of the panel in one direction and to the top of the box in the other. While the lower half of the rod "rode" back on the push-panel as it was opened, the upper half swung against a slender stick of wood which was attached to a light, plastic wheel mounted on a well-oiled axle. The initial force applied to the door was thus transmitted to the wheel and measured in degrees of rotation. Because the wheel offered very little resistance to the metal rod, the initial movement of the push-panel caused it to turn out of range of further movements of the metal rod. When the push-panel was opened approximately $\frac{1}{32}$ in., the clock stopped. In this manner the latency of the panel-pushing response was measured.

The food reward pellet was placed in a tray which was located approximately $\frac{1}{4}$ in. below the lower panel jamb. Water reward was administered at the same point through a curved tube from a water bottle which was retractable.

Habituation.—All *Ss* were handled for three days prior to the habituation phase of the experiment. Up to this time and during handling, unlimited food and water were available in the home cage. The *Ss* were divided into two major groups: a thirst group ($N = 12$) and a hunger group ($N = 24$). Each day for five days, *Ss* from both groups were placed individually in the activity chamber of the box for a 6-min. period, and their activity was recorded. On Day 3 the *Ss* received their initial deprivation. The hunger group was deprived of food for 23½ hr., and the thirst group was deprived of water for the same period. On Days 4 and 5, the hunger *Ss* were placed in individual feeding cages where, after a 6-min. delay, reward pellets were made available for 15 min. These pellets were followed by a wet mash composed of 60% water and 40% ground Purina dog chow by weight which was made available for an additional 15 min. Once or twice during this period, the mash was removed from the feeding cages and stirred. Given more time the *Ss* refused to eat, and so at the end of this 30-min. period all *Ss* were returned to their home cages where water was available for the next 23½ hr. but not food.

On Days 4 and 5, the thirsty *Ss* were placed in the individual feeding cages with Purina pellets immediately available and with water made available after 6 min. These *Ss* remained in the individual feeding cages for a period of 45 min. with water and dry food and were then returned to their home cages where food but no water was available for the next 23¼ hr.

Training.—Once per day, for the next three days, each *S* was placed individually in the activity chamber for a period of 6 min. All *Ss* at this time were under a 23½-hr. drive. At the close of this activity period, the push-panel was presented according to the following schedule: Day 1, door fully open for 8 trials and door ¼ in. open for 2 trials; Day 2, door fully open for 2 trials and door ¼ in. open for 8 trials; and Day 3, door ¼ in. open for 2 trials and door fully closed for 8 trials.

Hungry *Ss* were presented on each of these trials with reward pellets weighing .20 gm. Thirsty *Ss* were given access to the drinking tube for 5 sec. per trials (.20 cc.). Following each day's training, *Ss* were treated as on Days 4 and 5 of the habituation period.

Test period.—Each *S* was next given a total of 40 test trials, 20 successive trials (10 per day) at a low-drive level and 20 successive trials (10 per day) at a high-drive level. The blocks of high- and low-drive trials were counterbalanced. Half of each group was tested under high drive first and then shifted to low drive (H-L), and the other half was tested first under a low drive and shifted to a high (L-H).

In order to insure that *Ss* operating under a low drive would respond for the full 10 trials per day, it was necessary to determine empirically the minimum deprivation time for this to occur. Using animals other than the *Ss* of the present study, this value was determined to be 5 hr. for water deprivation and 8 hr. for food deprivation. It should be noted that this privation is essentially from an ad libitum diet. Hungry *Ss* under low drive were fed to satiation with water available every running day 8 hr. before testing and 6 min. after testing. Thirsty *Ss* under low drive were watered to satiation with food available every running day 5 hr. before testing and 6 min. after testing. High drive for hungry *Ss* was arbitrarily set at 48 hr. of food deprivation. These *Ss* were fed to satiation every running day 6 min. after the test trials. High drive for thirsty *Ss* was set at 36 hr. of water deprivation. These *Ss* were watered to satiation with food available 6 min. after test trials every running day, and then again at noon on the following day and run 36 hr. later. The satiation procedure was the same as that described for Days 4 and 5 of the habituation period. For reasons previously mentioned,[4] the hunger group was divided into two equal subgroups: a large reward group (.32 mg.) and a small food reward group (.08 gm.). All *Ss* were run between the hours of 8 P.M. and 1 A.M.

Results

The latency measure.—Figure 2 presents a plot of the mean latencies for

[4] See Footnote 3.

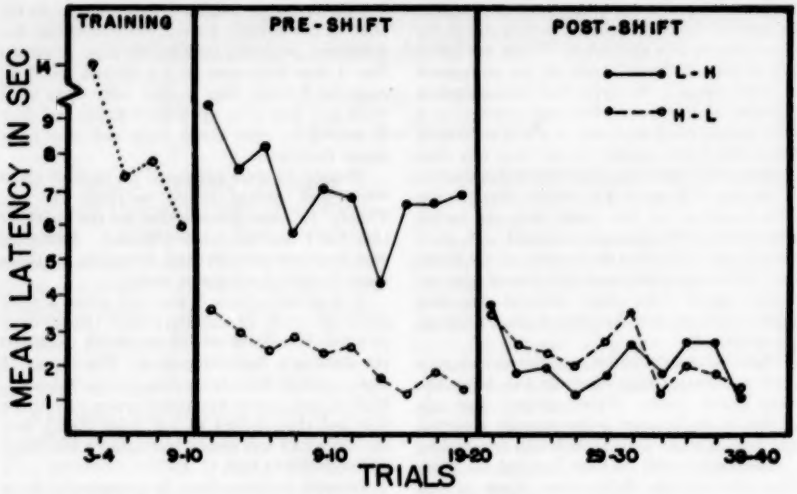


FIG. 2. Latency in seconds of H-L and L-H *Ss* during last 8 training trials and 20 preshift and 20 postshift trials.

the last 8 training trials and the 40 test trials (preshift and postshift) for all 36 *Ss*. In order to smooth the curves, latencies on Test Trials 1 and 2 through 39 and 40 were paired, and the mean latency of each pair was plotted. These test trial curves reveal a trend similar to that found by Deese and Carpenter (4). Animals shifted from low drive to high drive showed a marked decrease in latency, whereas *Ss* switched from high drive to low drive continued to perform at about the same level as obtained prior to drive reversal.

TABLE 1

MEAN LATENCY/TRIAL IN SECONDS FOR 20 HIGH-DRIVE AND 20 LOW-DRIVE TRIALS

(N = 6 each group)

Reinforcement	High Drive		Low Drive	
	Mean	*SD*	Mean	*SD*
Large Food				
H-L	2.65	1.56	2.61	.95
L-H	2.51	1.53	3.42	1.26
Small Food				
H-L	2.56	1.76	1.70	1.09
L-H	1.75	.66	7.39	2.81
Water				
H-L	1.64	.80	2.49	.84
L-H	2.44	1.26	8.90	3.68
Total	2.26	1.40	4.41	1.88

The mean latency scores for each of the reinforcement subgroups is presented in Table 1. In two of the three subgroups, latencies decreased rather than increased when *Ss* were shifted from high drive to low drive. The difference between the mean high-drive latency and the mean low-drive latency for each of the two drive shift conditions (H-L and L-H) was tested using a matched t. The mean high-drive latency of L-H *Ss* was significantly shorter than the mean low-drive latency of the same population (t = 5.8, $P < .01$). However, there was no significant difference between the mean high-drive latency and the mean low-drive latency of

H-L *Ss* ($t = .036$). The same results were obtained when a sign test was used for these comparisons.

The activity measure.—Figure 3 presents a plot of the mean activities of *Ss* in the two drive shift groups on each of the four test days. As in the case of the latency measure, a trend toward drive reversal was apparent. For each of the drive shift groups, a matched t was computed to compare activity when measured under high-drive and under low-drive conditions. A significant increase in activity was obtained when *Ss* were shifted from low drive to high drive ($t = 5.6$, $P < .01$), but when *Ss* were shifted from high drive to low drive, the decrease in activity was not significant ($t = .167$). When a composite curve (See Fig. 3) was drawn based on these data, a consistent rise in activity was obtained over the four test days. The mean activities measured under the various drive shift and reinforcement conditions are presented in Table 2.

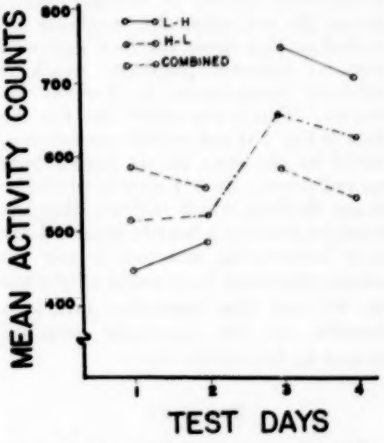


FIG. 3. Mean activity counts on each of the 4 experimental days of *Ss* in the 2 drive shift groups. Continuous curves based on combined data of both groups.

TABLE 2

MEAN ACTIVITY PER TEST DAY AT HIGH-DRIVE AND LOW-DRIVE LEVELS

($N = 6$ each group)

Reinforcement	High Drive		Low Drive	
	Mean	*SD*	Mean	*SD*
Large Food				
H-L	496	124	535	156
L-H	811	178	507	103
Small Food				
H-L	582	49	553	140
L-H	748	110	488	191
Water				
H-L	638	242	592	232
L-H	609	117	384	184
Total	647	170	509	184

The amplitude measure.—Obvious non-normality in the distribution of the amplitude scores required the application of a nonparametric statistic to the data. Mean high-drive amplitudes and mean low-drive amplitudes were compared using the sign test (6). When the drive shift procedure was ignored, mean high-drive amplitude scores exceeded mean low-drive amplitude scores in only 19 of the 36 cases, and the null hypothesis was accepted. Drive, in other words, did not significantly affect the amplitude of response. When these data were broken down in terms of the drive shift procedure employed, the null hypothesis was accepted in both instances. Mean high-drive scores exceeded mean low-drive scores in only 8 out of 18 H-L cases and 11 out of 18 L-H cases. Although both of these differences were in the direction which would be predicted on the basis of the previously reported drive-reversal data, neither was significant.

DISCUSSION

The drive-reversal effect was clearly obtained in the case of latency and

activity. However, since there was no difference between high- and low-drive amplitudes, the conditions for a drive-reversal effect did not even exist with respect to this measure. Although Hull has stated that response amplitude is one of the aspects of behavior which varies as a function of drive, the evidence for this is based primarily on autonomic measures (**7, 8**). Since the measures utilized in the present study were skeletal, the results do not directly contradict the evidence on which the assumption is based.

Within the original theoretical framework of this study, there are two possible explanations for the drive-reversal effect. One of these is the possibility of residual motivation. However, since the *S*s in the present study were completely satiated twice in the 24-hr. period following the shift from high drive to low drive, it appears unlikely that residual motivation could account for the continued high level of performance of *S*s shifted to low drive. Moreover, between Days 1 and 2, low-drive *S*s increased their mean activity count by 24 while high-drive *S*s showed a decrease in activity for the same period. If residual motivation were building up, we would expect it to reflect itself in a greater increase in activity in those *S*s undergoing the greatest deprivation which is not the case.

The second possible explanation for the drive-reversal effect and one which appears more adequate is based on Hull's secondary motivation corollary (**8**, p. 25). This corollary states that neutral stimuli, which are repeatedly and consistently associated with the evocation of a drive, acquire the capacity to bring about the drive stimuli. In the present study, the abrupt diminution in drive was introduced after a large number of trials had been administered. Presumably, many of the neutral stimuli in the box could be conditioned to the primary drive stimuli during these trials and these previously neutral stimuli could become the basis for the continued high level of performance of low-drive *S*s in the postshift phase of the study. The most convincing bit of evidence in favor of this explanation is to be found in Fig. 2. An inspection of this figure reveals that the unilateral drive-reversal effect did not occur when *S*s were shifted from an intermediate drive to either a high drive or a low drive following the training phase of the experiment. The fact that no drive-reversal effect was apparent at this point in the study and the fact that it became apparent 20 trials later, supports the hypothesis that the neutral stimuli were acquiring a motivating property during this period.

A third interpretation, which, theoretically, may stray a bit from the other two, should also be mentioned. It is possible to assume that the responses being conditioned under high drive differ from the responses being conditioned under low drive. Under high drive the panel-pushing (food approach) response is more consistently elicited and there are fewer competing responses occurring (**2**). Once this approach response is conditioned under high-drive conditions, it is tied to the cues of the situation and remains strong even when the drive is reduced. On the other hand, under low drive no response is being consistently elicited or strongly conditioned (**5**) and when the conditions are shifted to high drive the food approach response becomes prepotent which is reflected immediately in lowered latencies. This explanation also fits the data of Fig. 2 as the preshift conditioning would be the same for *S*s shifted both up and down. Such a view is related to recent findings which indicate that the habit for pressing a heavily weighted bar in a bar-pressing situation is not the same as the habit for pressing a light bar (**9, 10**) and that extinction rate is a function of the *particular* response elicited by frustration (**1**).

Summary

The drive-reversal effect, i.e., the tendency of animals shifted from high to low drive to continue performing at the same level as obtained prior to drive shift, was investigated with respect to three performance measures—latency,

activity, and response amplitude. Thirty-six *S*s were trained in a panel-pushing apparatus at an intermediate drive level. The *S*s were then given 20 trials at a high-drive level and 20 trials at a low-drive level in accordance with one of two drive-shift procedures: H-L or L-H. A drive-reversal effect was obtained in the case of latency and activity but not in the case of response amplitude. Several alternative explanations for this effect were discussed.

References

1. Adelman, H. M., & Maatsch, J. L. Resistance to extinction as a function of the type of response elicited by frustration. *J. exp. Psychol.*, 1955, **50**, 61–65.
2. Cotton, J. W. Running time as a function of food deprivation. *J. exp. Psychol.*, 1953, **46**, 188–198.
3. Davis, R. H. The quantification of drive: I. Incentive values of food and water. Unpublished doctoral dissertation, Michigan State Univer., 1953.
4. Deese, J., & Carpenter, J. A. Drive level and reinforcement. *J. exp. Psychol.*, 1951, **42**, 236–238.
5. Denny, M. R., & Adelman, H. M. Elicitation theory: I. An analysis of two typical learning situations. *Psychol. Rev.*, 1955, **62**, 290–296.
6. Dixon, W. J., & Massey, F. J. *Introduction to statistical analysis.* New York: McGraw-Hill, 1951.
7. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
8. Hull, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.
9. Maatsch, J. L., Adelman, H. M., & Denny, M. R. Effort and resistance to extinction of the bar-pressing response. *J. comp. physiol. Psychol.*, 1954, **47**, 47–50.
10. Stanley, W. C., & Aamodt, M. S. Force of responding during extinction as a function of force requirement during conditioning. *J. comp. physiol. Psychol.*, 1954, **47**, 462–464.

(Received April 26, 1956)

Journal of Experimental Psychology
Vol. 53, No. 5, 1957

RESPONSE STRENGTH AS A FUNCTION OF DELAY OF REWARD IN A RUNWAY[1]

WAYNE B. HOLDER,[2] MELVIN H. MARX, ELAINE E. HOLDER,[2] AND GEORGE COLLIER

University of Missouri

A number of experiments have investigated the effect during acquisition of delay of reinforcement on response strength. In general the results have indicated that both the rate and asymptote of learning are some decreasing function of the length of delay and of the distance of the delay from the reward.

A second, motivational, effect is sometimes ascribed to delays of reinforcement occurring in response chains having some strength. For example, Amsel (**1, 2**) and Brown and Farber (**4**) have suggested that delay of reinforcement is one of the conditions that produce frustration. Such frustration is assumed to have the functional status of an irrelevant drive and to strengthen any response which accompanies or follows the frustrating situation by increasing the general motivational level of the organism. It is also assumed that the frustration-aroused drive has typical drive-stimulus properties, and further that any stimulus which has been associated with the frustration situation acquires conditioned motivational properties which will operate to increase response strength in other situations. It has been suggested that the amount of frustration produced is some function of the length of delay.

Most of the preceding studies of delay have examined its effects when it closely preceded the goal response, and have not employed post-delay measures of response strength.

The present study examines, following acquisition, the changes in response strength before and after delay as a function of the length of delay, and the effect of a stimulus associated with delay of reward on the acquisition of response strength in a different situation. It was designed to test some of the implications of the Brown and Farber (**4**) treatment of frustration.

METHOD

Apparatus.—The first runway consisted of a 90-in. unpainted, straight alley 5½ in. deep and 3 in. wide; the length included a 12-in. starting box, a 12-in. goal box, and a 14-in. delay box halfway between the starting box and the goal box. Natural wood doors were at both ends of the delay box and there was a Plexiglas door in front of the second delay box door. When the delay box doors were raised the delay box was not distinguishable in the runway. The starting box and goal box doors were dark brown tempered masonite.

A contact switch on the starting box door started a Standard Electric timer when the door was opened. When *S* broke a photoelectric beam 6 in. from the starting box door, this timer stopped and a second timer started. The second timer was stopped in the same way immediately in front of the first delay box door. A third timer was started by opening the door leading out of the delay box and stopped when *S* crossed a photoelectric beam 6 in. in front of the delay box door. A fourth timer was started by this photoelectric cell and stopped when *S* broke a beam of light immediately in front of the goal

[1] This investigation was supported in part by a research grant, M-817, from the National Institute of Mental Health, of the National Institutes of Health, Bethesda, Maryland. This experiment was planned and executed by the first author under the general supervision of the second with assistance from the third; the fourth author is primarily responsible for the statistical analysis and the form of the present report.

[2] Now at Fresno State College, California.

box door. Both runway segments were 26 in. in length.

The second runway consisted of a 72-in. straight alley 5½ in. deep and 3 in. wide; the length included a 12-in. starting box and a 12-in. goal box. This entire apparatus was painted a flat black.

Subjects.—Thirty-nine male albino rats were used. They were approximately 120 days old at the beginning of the experiment, at which time they were placed on a 23-hr. feeding schedule for 10 days and habituated to *E* and the apparatus.

Procedure.—Training trials were given in the first runway. All *S*s were given one training trial per day for 25 days with approximately 1 sec. delay and a .3-gm. food reward. The Plexiglas door was down on these trials, forcing *S* to stop, and was raised after 1-sec.

Ten test trials, one per day, were then given in the same runway, with .3 gm. food reward, under the following conditions: 1-sec. group, 1 sec. delay with the same procedure as used in the training trials; 15-sec. group, 15 sec. delay; and 45-sec. group, 45 sec. delay. The wooden doors of the delay box were lowered for the 15-sec. and 45-sec. delay groups. A small muffled doorbell buzzer sounded during the entire delay interval for the 15-sec. and 45-sec. groups. At the termination of delay, the wooden doors were raised first, followed immediately by the raising of the Plexiglas door. Starting times (ST) and running times (RT) preceding and following the delay box were measured on all training and test trials.

One test trial per day was given for 10 days in the second runway. The buzzer from the first series of test trials was mounted on the starting box of this runway and was sounded throughout the trials for all three groups.

Results

First runway (delay of reward).—An analysis of variance was performed on Log 1/ST[3] on the last five training trials for delay group (D), trials (T), and the first and second runway segment (R). T and R proved to be significant, $P(F_T = 3.148;\ 4,\ 144) < .025$ and $P(F_R = 92.076;\ 1,\ 36) < .001$, respectively. These results indicate that with respect to starting times the rats had failed to reach asymptote and that the starting times from the start box were significantly shorter than those from the door of what was to become the delay box. The data of this analysis are presented in the first and third panels of Fig. 1.

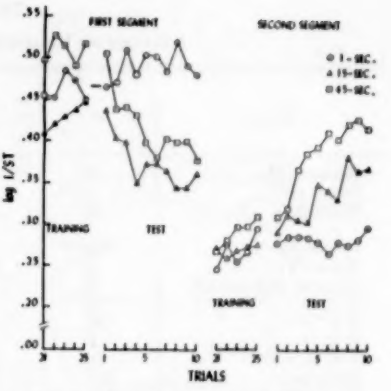


Fig. 1. First Runway: Log 1/ST.

An analysis of variance on Log 1/ST for length of delay (D), trials (T), and runway segment before and after delay (R) is presented in Table 1. The main effect of R and the interactions of R × T and R × D were significant. The data of this analysis are presented in the second and fourth panels of Fig. 1. Tests of trends for the individual T × R × D combinations were performed using the appropriate error term from the over-all analysis. Both the before

[3] Log (RT + 1) and Log 1/ST yielded the most symmetrical distributions and the most homogeneous variances of the measures examined and were therefore used in the analyses. The first measure increases, the second decreases, with response strength. The results of the same analyses on the raw data, RT and ST, and on the other transforms tried, yielded in every case the same conclusions, with one very minor exception. Since the major comparisons among the curves in Fig. 1 and 2 differ significantly nonparametrically for both the transformed and untransformed data, the results of the analysis of variance are not artifacts of the transformations. The results were, therefore, summarized in terms of RT and ST, both of which decrease with response strength.

TABLE 1

ANALYSIS OF VARIANCE FOR THE FIRST RUNWAY

Source	*df*	Log 1/ST		Log (RT + 1)	
		MS	*F*	*MS*	*F*
Between *S*'s	38				
Delay (D)	2	.1422	.852	.00227	.260
Error (b)	36	.1668		.00865	
Within *S*s	664				
Trials (T)	8	.0030	.946	.000157	.860
Runway segment (R)	1	1.2373	27.573***	.02468	21.344***
T × R	8	.0183	2.819**	.00072	3.928***
T × D	16	.0025	.817	.00012	.687
R × D	2	.7294	16.253***	.03804	32.896***
T × R × D	16	.0090	1.374	.00032	1.795*
Error (w)	612	.0072		.00024	
$Error_1$(w)	288	.0031		.00018	
$Error_2$(w)	36	.0449		.00116	
$Error_3$(w)	288	.0065		.00018	
Total	701				

* $P < .025$.
** $P < .005$.
*** $P < .001$.

and after delay curves for 15-sec. and 45-sec. delays showed significant trends at least at the .05 level, while neither of the 1-sec. curves showed a significant trend. From the plots and the analysis it can be seen that ST's from the start box increased over trials as a function of length of delay and ST's from the delay box decreased over trials with length of delay. For the before-delay ST's there is an inversion between the 15-sec. and 45-sec. group, the former showing the longest ST's. This inversion is nonsignificant by the analysis of variance, although by the binominal expansion the curves differ at the .01 level. Whether this inversion was due to the initial, nonsignificant, difference or represents an experimental difference could not be determined from these data. The failure of either of the 1-sec. curves to show any significant trend indicates that the ST's were near asymptote. The initial difference in ST's for the two runway segments appearing during training remained during test.

An analysis of variance was performed on Log (RT + 1) on the last five training trials for delay group (D), trials (T), and first and second runway segment (R). Only R proved to be significant, $P(F_R = 29.189;\ 1, 144) < .001$, indicating that *S*s were near asymptote on this measure and that they ran the second runway

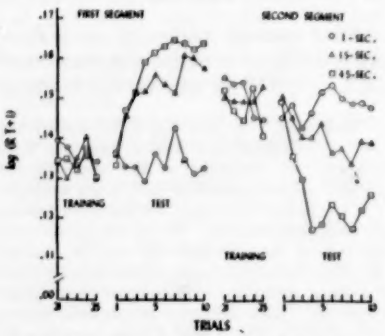


FIG. 2. First Runway: Log (RT + 1).

segment significantly more slowly. The data of this analysis are presented in the first and third panels of Fig. 2.

An analysis of variance on Log (RT + 1) for length of delay (D), trials (T), and runway segments before and after delay (R) is presented in Table 1. The main effect of R, and the interactions of R and T, R and D, and R, T and D, were significant. The data of this analysis are presented in the second and fourth panels of Fig. 2. Tests of trends for the individual T × R × D combinations were performed using the appropriate error term from the over-all analysis. Both the before- and after-delay curves for the 45-sec. group and the before-delay curve for the 15-sec. group showed significant trends, while neither the 1-sec. curves nor the curve for the 15-sec. group following delay showed any significant trends. The trend of the 15-sec. curve, following delay, was very nearly significant at the .05 level. The over-all differences between the 15-sec. and 45-sec. groups and the 15-sec. and 1-sec. groups for before and after delay separately were significant at least at the .05 level. From the plots and the analysis it can be seen that RT's from the start box increased as a function of length of delay over trials, and RT's from the delay box decreased as a function of length of delay over trials. In both cases the rate of increase or decrease was also a function of length of delay. It may further be noted that while there was a significant over-all difference in RT's between the two runway segments, the 45-sec. group following delay reached a level of RT's significantly faster than that for all other segments and groups (cf. Fig. 2).

The correlations between Log 1/ST and Log (RT + 1) for Trial 2 and for Trial 10 before delay were − .738 and − .740, respectively. The same correlations after delay were − .562 and − .751. There were no significant group differences for the correlations computed separately. Linearity and homoscedasticity were not rejected.

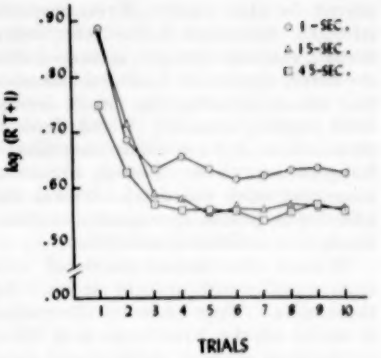


FIG. 3. Second Runway: Log (RT + 1).

Second Runway.—An analysis of variance was performed on Log (RT + 1) for delay group (D) and trials (T). Both T and D proved to be significant while the T × D interaction did not, $P(F_D = 4.589; 2, 32) < .0025$, $P(F_T = 37.738; 9, 324) < .001$, and $P(F_{TD} = 1.193; 18, 324) > .05$. The data of this analysis are presented in Fig. 3. The difference between the 15- and 45-sec. delay curves was separately tested and failed to reach significance. These plots and the analysis show that the RT during the course of training in the second runway is some function of the length of delay on the first runway. For the number of trials run, the curves do not become significantly nonparallel.

DISCUSSION

At least four very interesting effects of delay introduced into a chain of responses at nearly asymptotic strength were ap-

parent in this study. First, response strength decreased before the delay. Second, response strength increased after the delay, apparently to a level exceeding that obtained during the nearly asymptotic training sessions. Third, both of these effects did not occur immediately but grew over trials. Fourth, a stimulus associated with the delay interval was effective in producing a greater response strength in a different situation.

At least two current points of view seem equally well able to account for these data. They may be interpreted in terms of the hypothesis that delay results in an aversive, motivational state, whose magnitude is a function of the length of the delay, and which can be evoked by stimuli associated with the conditions producing the state (1). On this hypothesis the decrease in response strength in the segment preceding delay is either the result of the learning of an avoidance response to the noxious stimuli associated with frustration (4) or the development of fractional anticipatory frustration response with inhibitory properties (1). The increased response strength following the delay would be accounted for in terms of the irrelevant-drive properties of frustration. The increased response strength in the second runway would be the result of the conditioning of the irrelevant motivation aroused by the delay. This hypothesis differs from the usual accounts (e.g., 11, 19, 20) of the effects of delay which predicate a decline in response strength as a consequence of increased delay of reward. Hull (11) accounts for this in terms of a motivational variable, *J*, which multiplies habit strength. Spence (20) offers an alternative account relating *J* to the magnitude of the conditioned fractional anticipatory goal response and giving it associative properties.

The data may also be interpreted from an associative point of view. According to this hypothesis, the decrease in response strength in the runway segment preceding delay is the product of extinction resulting from the increased delay in reward; that is, the response strength falls to a new asymptote commensurate with the delay of reward. The increase of response strength in the runway segment following delay would be the joint product of the adaptation or extinction of competing responses to the block or the delay box door, the growth of strength of orientative and preparatory responses, similar to those in the start box, and the closer proximity of the reward in the second segment. Thus the effect of the delay would be functionally to cut the runway in two with the consequent shallower gradient of reinforcement (12) and faster RT in the second segment. The greater response strength shown by the delay groups in the second runway is due to the buzzer becoming a conditioned stimulus for "faster" running.

The data of the first segment seem to be completely impartial. Avoidance learning (4), fractional anticipatory frustration (1), or extinction seem equally able to handle the results. Since the start box was farthest from reward, start-box ST would be expected, on ordinary delay-of-reinforcement notions, to have gained the least response strength, as it apparently did; and therefore, to be least affected by the introduction of delay. If this reasoning is correct, the inversion of the 15-sec. and 45-sec. group on ST from the start box is not disturbing to any of the points of view, since it would, then, be most sensitive to the operation of extraneous factors, such as original group differences, which might have been operating. All three hypotheses would predict that both ST and RT would fall to new, lower asymptotes of response strength. The associative account is similar to the one given by Brown and Gentry (5, 6) for the failure of their delay rats to learn the pre-delay segment of their maze and Holder (10) in accounting for the increased RT of rats shifted from 40- to 70- and 100-sec. delays. Cooper (7) and Gilhousen (9) have reported similar decreases in RT for rats delayed in the middle of a maze.

While there is rough qualitative agreement, the data of the second segment pose a serious difficulty for the irrelevant-

motivation hypothesis. A number of studies have demonstrated, in varied situations, that increased vigor of response follows interruption of, delay of, or failure to reward. For example, Marzocco (14), Rohrer (16), and Sheffield (18) present data in which the effect of failure or delay of reward is presumed to manifest itself immediately. On the other hand, Geier and Tolman (8), who obtained measures of activity during delay, showed an increasing effect (on activity) over trials, while Logan (13) showed a gradual decline of response strength over trials as the result of the introduction of an increased delay. This growth of response strength over trials presents the major difficulty for the irrelevant-drive hypothesis. The effect of blocking should result, following Amsel (2) or Brown and Farber (4), in an immediate increment in drive which should multiply whatever habits are present and produce an immediate increment in response strength. Hull's latest formulation (11) would predict that the introduction of delay would result in an immediate decrement in response strength. The spacing of the test trials in the present experiment precludes any interpretation in terms of an accumulation of motivation or of any immediate increment, and necessitates some sort of associative process with respect to the irrelevant motive in order to account for the gradual change in response strength. Amsel (1) has suggested one such mechanism in his fractional anticipatory frustration concept. A number of other such explanations could be invented (e.g. 17).

The inverse relation between response strength and length of delay, with the 45-sec. group's RT's falling below those of any other condition, presents the major difficulty for the associative account. The effect of interrupting a run is shown by all groups during training where both ST and RT are consistently slower following delay. This may result, on the present hypothesis, from unconditioned (emotional) or conditioned responses to blocking or to "irrelevant" responses to the new stimuli. Where these responses compete with the ongoing behavior, a decrement would be expected, as in the present study; where they do not, facilitation would be expected as, for example, in Amsel and Roussel (2) or Marzocco (14). Thus the differential effects of delay would be accounted for in terms of the differential amounts of adaptation or extinction of competing responses during delay and the greater likelihood of a "well-oriented" rat. Some support for this notion is provided by the results of Brown and Gentry (5, 6), who found no differences in RT in the portion of a maze following delay between a control group (not delayed) and a group delayed for 45 secs. on a delay wheel. The results of Amsel and Roussel (2) can also be accounted for in this way. Since they failed to report the RT's and ST's for the first segment of their maze and since they used each rat as its own control, alternating reward and nonreward conditions with the resulting possibility of confounding and obscuring the goal gradient and like effects, it is difficult to give a clear, *post hoc* account of their results.

Resistance to extinction measures, which were not possible here, would supply useful information with respect to the covariation of RT and ST with number of trials to extinction. As is suggested in the Sheffield (18) and Stanley (21) studies, little relation would be expected from the irrelevant motivation point of view; that is, RT and ST would presumably be sensitive to increased motivation, while habit strength as measured by trials to extinction, would not. The opposite prediction would be made by the associative hypothesis. Information on the operation of secondary rewards in the present situation would also be useful in making a decision between the two hypotheses.

One problem posed for both accounts is the role of the buzzer used as a CS. While the dampened buzzer was practically imperceptible to humans, it is known that white noise at 120 db. is noxious for the albino rat (15). The present results might be accounted for in terms of "time in the presence of the

noxious stimulus." The fact that the 1-sec. group, which had not previously been exposed to the buzzer, was the slowest group in the second runway, and that gradual learning curves occurred for all groups, argues against this interpretation of the results.

Both hypotheses would predict that response strength in the second runway would be some increasing function of the length of delay. The present study and that of Bernstein (3) are the only two which have attempted to condition the effects of delay of reward.

In both studies the results may be interpreted as the conditioning of some motivational state or as the conditioning of responses or components of responses used in the second situation. The latter is more difficult in the Bernstein study since there was no mediating conditioned stimulus. The failure of the 15-sec. and 45-sec. groups to differ in the present experiment could be accounted for by either hypothesis in terms of the lability of the conditioning which occurred in the small number of pairings.

While the present data have been indecisive with regard to theory, their clarity sheds much light on the effect of delay in the runway situation and offers a number of interesting leads for further research.

Summary

Thirty-nine albino rats were trained to near asymptote on a straight runway in which each run was interrupted in the middle of the runway for 1 sec. Following training, *S*s were divided into three groups. The first group continued under the training conditions, the second was delayed in the middle of the runway for 15 sec., and the third for 45 sec. A buzzer was sounded during the delay period for the 15-sec. and 45-sec. groups. The three groups were then run on a second runway in which the buzzer which had been associated with delay was sounded throughout a run.

On the first runway it was found that response strength, as measured both by ST and RT, was a decreasing function of the length of delay in the segment of the runway preceding the delay, and an increasing function of delay in the segment of the runway following delay. On the second runway the acquisition curves for RT for both of the delay groups fell below that for the control.

Alternative interpretations of the data were outlined in terms of two sets of hypotheses: (*a*) that delay of reinforcement occurring in a chain of highly trained responses produces an aversive motivational state which can be evoked by a stimulus associated with conditions producing the state and which is added to the ongoing motivational complex; and (*b*) that delay of reinforcement extinguishes the members preceding it in a response chain, that the increased vigor of response in the second segment is the joint result of differential extinction and adaptation of competing responses to the delay, and the closer proximity of the reward, and that stimuli associated with these responses will tend to evoke similar responses in similar situations.

References

1. Amsel, A. A three-factor theory of inhibition: an addition to Hull's two-factor theory. Paper read at the Southern Society for Philosophy and Psychology, Roanoke, Va., March, 1951.
2. Amsel, A., & Roussel, J. Motivational properties of frustration: I. Effects on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, **43**, 363–368.
3. Bernstein, B. B. Extinction as a function of frustration. Unpublished doctor's dissertation, Univer. of Missouri, 1954.
4. Brown, J. S., & Farber, I. E. Emotions conceptualized as intervening variables with suggestions toward a tentative theory of frustration. *Psychol. Bull.*, 1951, **48**, 465–495.
5. Brown, W. L., Gentry, G., & Kaplan, S. J. The effect of intra-maze delay. I. Delay enforced by a revolving wheel. *J. comp. physiol. Psychol.*, 1948, **41**, 258–268.
6. Brown, W. L., & Gentry, G. The effect of intra-maze delay. II. Various intervals of delay. *J. comp. physiol. Psychol.*, 1948, **41**, 403–407.
7. Cooper, J. B. The effect upon performance of introduction and removal of a delay within the maze. *J. comp. Psychol.*, 1938, **26**, 457–462.
8. Geier, F. M., & Tolman, E. C. Goal distance and restless activity. I. The goal gradient of restless activity. *J. comp. Psychol.*, 1943, **35**, 197–204.
9. Gilhousen, H. C. Temporal relations in anticipatory reactions of the white rat. *J. comp. Psychol.*, 1938, **26**, 163–175.
10. Holder, E. Response strength as a function of changes in the length of the delay interval. Unpublished master's dissertation, New Mexico State College, 1951.

11. HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
12. HULL, C. L. The rat's speed-of-locomotion gradient in the approach to food. *J. comp. Psychol.*, 1934, **17,** 393–422.
13. LOGAN, R. A. The role of delay of reinforcement in determining reaction potential. *J. exp. Psychol.*, 1952, **43,** 393–399.
14. MARZOCCO, F. N. Frustration effect as a function of drive level, habit strength and distribution of trials during extinction. Unpublished doctor's dissertation, State Univer. of Iowa, 1950.
15. POLAND, R. G., HELSTROM, E. A., & DAVIS, R. T. Running speed of rats as a function of auditory stimuli. *Proc. S. D. Acad. Sci.*, 1952, **31,** 177–181.
16. ROHRER, J. H. A motivational state resulting from non-reward. *J. comp. physiol. Psychol.*, 1949, **42,** 476–485.
17. SEWARD, J. P. Secondary reinforcement as tertiary motivation: a revision of Hull's revision. *Psychol. Rev.*, 1950, **57,** 362–374.
18. SHEFFIELD, V. F. Resistance to extinction as a function of the spacing of extinction trials. *J. exp. Psychol.*, 1950, **40,** 305–313.
19. SPENCE, K. W. The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, 1947, **54,** 1–8.
20. SPENCE, K. W. Theoretical interpretations of learning. In C. P. Stone (Ed.), *Comparative psychology.* New York: Prentice-Hall, 1951.
21. STANLEY, W. C. Extinction as a function of the spacing of extinction trials. *J. exp. Psychol.*, 1952, **43,** 249–300.

(Received April 27, 1956)

Journal of Experimental Psychology
Vol. 53, No. 5, 1957

FACTORS CONTRIBUTING TO GENERAL VERSUS SPECIFIC PERCEPTUAL LEARNING[1]

J. ALFRED LEONARD,[2] H. WESTON CLARKE, AND SARA R. STAATS

The Ohio State University

Harlow's studies on the formation of learning sets (4) have revived interest in the "increments of performance that occur as a subject passes from sample to sample of a single class of material" (6, p. 307). In general, it has been found that experience with and training on one sample of a class facilitates learning on another sample of the same class.

We became interested in these findings for two separate though related reasons. First, although the phenomenon is one of the best established in learning literature no attempts seem to have been made to analyze it in detail. For instance, transfer between samples has usually been measured by reference to the saving relative to initial performance, but not with reference to the performance of a nontransferring control group at comparable posttransfer stages. Similarly, it is not clear to what extent the transfer effects found can be ascribed directly to familiarity with the class of material as apart from familiarity with the task or experimental situation. It seemed possible to answer such questions by a suitable manipulation of the pretraining conditions of separate groups of *Ss* and comparing the effects of such groups on transfer to a common situation.

Second, it seemed feasible to utilize the information so obtained for the purpose of making inferences about the relative importance of various aspects of the total learning situation to the performance of the nontransferring control group. The development of stimulus material and methods for the study of visual figure recognition described by Fitts *et al.* (2) provided an opportunity for investigating some of the problems raised. The complexity levels of their "metric figures" are defined by the number of stimuli in a specifiable population. At the higher levels of complexity it is possible to generate very large numbers of figures, and it is convenient to make use of several samples of limited size for experimental purposes. Moreover, one has a priori grounds for assuming equal difficulty among samples drawn at random from the same population provided the samples are of adequate size.

Considering a group of *Ss* operating with a sample of metric figures on a standard identification task one can attribute the level of performance at any given stage during training to learning of the properties of the class of stimuli, the population of stimuli, the specific figures making up the sample, and the task required. In the experiment to be described, the pretraining conditions of four groups were manipulated so as to include all or some of these components of the

[1] This research was conducted in the Laboratory of Aviation Psychology of The Ohio State University through support from a contract between the Office of Naval Research and the Ohio State University Research Foundation (Contract No. Nonr-495(02). The authors wish to record their indebtedness to Drs. G. E. Briggs, P. M. Fitts, and D. D. Wickens for encouragement and assistance.

[2] Visiting Research Associate from the Medical Research Council, Applied Psychology Research Unit, Cambridge, England.

common posttransfer condition: an identification task with samples of metric figures.

METHOD

Stimuli.—Two kinds of stimuli were used in this experiment. Properties of the metric figures have been described elsewhere in detail (2, p. 2). They are black on white, histogram-like figures surrounded by a .4-in. white border and mounted in 1.6 × 1.6-in. plastic covers. The present figures were ones in which any one of eight column heights could appear in any one of eight adjacent positions, i.e., the 8 × 8 random population as specified. Three samples of eight figures each were drawn from the total possible population of almost seventeen million figures. The other class of stimuli was easily distinguishable, black-on-white silhouettes of common objects (ship, cup, dagger, etc.) mounted on the same sized covers as those used for the metric figures.

Tasks.—Two tasks were used. The first was the standard identification task described in Fitts, *et al.* (2, p. 6). In this task *S*s are presented with a board on which a given sample of figures is replicated in six horizontal rows. The second was a classification task in which *S*s are presented with a pile of figures consisting of several copies of each figure making up the sample. One copy of each figure is laid out in front of *S* who has to pick up one figure at a time from the pile and place it in a stack below the appropriate test figure.

Experimental design.—This is shown in detail in Table 1. The four main groups differed among themselves in the conditions of pretraining but had in common the posttransfer condition. Each group of 45 *S*s was made up of three subgroups of equal size to permit the use of three samples of metric figures when required. Group I, acting as a control, had pretraining conditions identical with those obtaining during posttransfer. The *S*s in each of the three subgroups of Group II kept the identification task (I Task) constant, but transferred from one sample of metric figures to another. On transfer, therefore, the properties of specific figures were new to them. Group III kept samples of metric figures constant throughout, but were pretrained on the classification task (C Task). On transfer they had to learn the use of their figures in the standard identification task. Group IV again kept the standard identification task constant, but were pretrained on a sample of very easily distinguishable figures. They were as unfamiliar with metric figures when transferring to them as the members of all other groups at their first pretraining trial.

TABLE 1

EXPERIMENTAL DESIGN
(15 *S*s in each subgroup)

Group	Subgroup	Pretraining Trials 1 2 3	Posttransfer Trials 4 5 6
I	1	Metric A I Task	Metric A I Task
	2	Metric B I Task	Metric B I Task
	3	Metric C I Task	Metric C I Task
II	1	Metric C I Task	Metric A I Task
	2	Metric A I Task	Metric B I Task
	3	Metric B I Task	Metric C I Task
III	1	Metric A C Task	Metric A I Task
	2	Metric B C Task	Metric B I Task
	3	Metric C C Task	Metric C I Task
IV	1	Pictures I Task	Metric A I Task
	2	Pictures I Task	Metric B I Task
	3	Pictures I Task	Metric C I Task

An attempt was made to match, in terms of means and variances, those pairs of subgroups of Groups I and II which performed on identical samples during pretraining. Groups I and II were run first, followed by Groups III and IV.

Subjects.—The 180 *S*s were male and female volunteers from introductory psychology courses at The Ohio State University.

Procedure.—For the identification task this was as follows: the six replications of a given sample of figures were arranged on the sorting board, one to each horizontal row, in such a manner that the same figure would never appear above or below itself. From a seventh replication *E* selected one of the figures, placed it in the center above the top row of figures, and allowed *S* to inspect this figure for approximately 2 sec. before saying "Go" and starting the timing device. The *S* immediately began to look for the appropriate figure in the first row, turned it over when found, and proceeded in this manner through the sixth row. Having found the correct figure in that row, *S* had to press a button to stop the timing device. The *S* was allowed to use only one hand and had to find the correct figure in each row before being allowed to proceed. An easily distinguishable color code on the back of each figure provided the necessary knowledge of results. The *E*

recorded the time taken for this item, i.e., to find the six copies of the test figure on the board, while *S* replaced the turned-over figures. The *E* then selected another figure and repeated the above process until all the eight figures had been dealt with once, which was held to constitute one trial. The *E* then told *S* the time taken for this trial, i.e., the cumulative total for eight items. This time was accrued on a counter pulsed at the rate of 10 per sec., the clicks being clearly audible to *S* from the moment *E* said "Go" until *S* pressed the button. The figures in each row were then picked up and reshuffled before the whole process was repeated.

For the classification task the procedure differed primarily in that *S* was required to deal with the whole sample of figures without any interruption. He was again given the time per trial from the counter at the end of each trial.

The *S* stood in front of the sorting board with *E* standing beside him. Under all conditions, regardless of stimuli or task used, *Ss* had to make 48 responses per trial. The intervals between trials were of approximately 2 min. duration except for the one immediately preceding the transfer stage which lasted for 3 min. and during which *Ss* left the experimental booth. Instructions stressed both speed and accuracy.

Scoring.—The time per item (c.f., procedure above) was recorded to the nearest .01 sec. from a Standard Electric timer. Summed over the eight items constituting a trial this provided a time per trial. These values were divided by 48 (6 rows × 8 figures per sample) to give a mean time per figure per row per trial for each *S*. The results are presented in these terms in order to maintain uniformity with other publications in the series of studies on visual figure recognition at Ohio State University.

TABLE 2

MEAN TIMES PER FIGURE IN SECONDS

Groups		Pretraining Trials			Posttransfer Trials		
		1	2	3	4	5	6
I	1	3.43	2.72	2.46	2.43	2.21	2.12
	2				2.24	2.09	2.02
	3				2.30	2.22	2.14
II	1	3.40	2.72	2.46	2.96	2.63	2.41
	2				2.90	2.55	2.39
	3				2.44	2.10	2.01
III	1	4.01	3.41	3.06	3.08	2.66	2.43
	2				2.82	2.50	2.17
	3				2.75	2.38	2.17
IV	1	1.28	1.06	1.00	3.01	3.53	2.34
	2				3.09	2.63	2.42
	3				3.05	2.58	2.32

Note.—For Trials 1–3 the means are for 45 *Ss* each; for Trials 4–6, 15 *Ss* each.

RESULTS

The results are presented in Table 2 in terms of mean times per figure for the six trials. For the pretraining stage the data have been combined for the three subgroups within each main group, these data serving mainly to indicate the level at which pretraining took place. In evaluating the performance of Group III it should be recalled that this was the only group which did not perform the standard identification task during pretraining, but used the classification task. Although *Ss* in both tasks were required to make the same number of responses per task, the nature of those responses was different so that one cannot make a strict comparison between the two tasks. The case is entirely different, however, for Group IV, which performed the same task as Groups I and II but used very different stimulus material, chosen for its great simplicity. The difference between Groups I and II on the one hand and Group IV on the other is a direct measure of the relative difficulty of the stimulus materials used.

The results of an analysis of variance for the posttransfer data using a partially hierarchical model (**5,** p. 102) are summarized in Table 3. The attempted matching of pairs of subgroups belonging to Groups I and II was not taken into account, in view of the large number of degrees of freedom involved in the total analysis. The results indicate that there were significant effects due to differences in pretraining and to continued practice, and that there was significant interaction between these two effects. Any differences in performance due to samples of metric figures were shown to be nonsignificant.

TABLE 3

ANALYSIS OF VARIANCE FOR POSTTRANSFER DATA

Source	*df*	*MS*	*F*
Groups (G)	3	5.3435	7.30**
Samples (Sa)	2	1.4446	1.97
G × Sa	6	1.0168	1.39
Ss/G/Sa	168	.7321	
Trials (T)	2	12.1453	148.11**
G × T	6	.4927	6.01**
Sa × T	4	.0344	
G × Sa × T	12	.0127	
T × Ss/G/Sa	336	.0820	
Total	539		

** $P > .01$.

Tests of significance among means for the four main groups at each of the three trials were carried out using Duncan's method (1). In view of the large number of degrees of freedom it was decided to test for differences only at the 1% level. The appropriate error term for the test was T × Sx/C/Sa. Groups II, III, and IV all differed significantly from Group I for all three trials, and Group IV differed from II and III only on Trial 4, i.e., the first posttransfer trial. It was also shown that the significance of differences due to continued practice was a function of the pretraining conditions of the main groups; these differences were significant throughout for Groups III and IV only. For Group II there were significant differences between Trials 4 and 5, and 4 and 6, for Group I only between Trials 4 and 6.

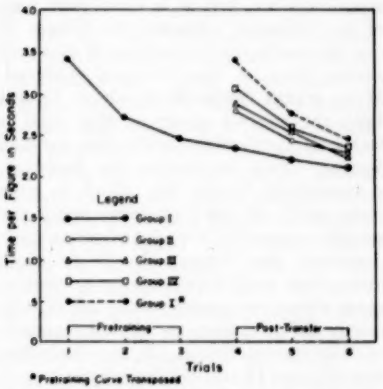


FIG. 1. Pretraining and posttransfer learning curve for Group I, posttransfer curves only for the other groups. The pretraining curve for Group I has also been transposed to the posttransfer position for purposes of comparison.

In Fig. 1 learning curves are presented for the pretraining trials of Group I and the posttransfer trials of all four groups. To enable a comparison of all the variations in pretraining used, the data of Group I for Trials 1–3 have been transposed to Trials 4–6, indicating the effect of no pretraining at all.

For present purposes the most meaningful measure of transfer is the one which relates performance of a given experimental group to both the initial performance of the control group as well as the control group's performance at the same posttransfer trial. In Gagné's terminology this is really an improvement score and is calculated from the formula

$$\frac{C_1 - E_x}{C_1 - C_x} \times 100,$$

where C stands for control group, E for experimental group, subscript 1 for trial 1, and subscript x for a given posttransfer trial (**3**, p. 102, formula 4). This measure takes the control group's improvement between Trials 1 and x as 100% and expresses the experimental group's performance on Trial x in relation to this. In Table 4 these values are given for Groups II, III, and IV for Trials 4 and 6.

DISCUSSION

Considering Groups I and II by themselves it has been shown that for the

TABLE 4

TRANSFER PERCENTAGES FOR GROUPS II, III, AND IV FOR TRIALS 4 AND 6

Group	Trial 4	Trial 6
II	59.45%	86.56%
III	48.64%	87.31%
IV	34.23%	79.85%

present type of material, transfer among different samples of the same class of material resulted in about 60% transfer on the first posttransfer trial, increasing to about 87% on the third posttransfer trial. Both groups continued to improve performance between Trials 4–6, but clearly Group II had the greater improvement so that amount of transfer increased between Trials 4–6. A comparison of the data of Groups I and IV permits amplification of the statement concerning the efficiency of transfer among samples of material. It seems fair to say that Group IV operated under conditions most nearly similar to those used in more conventional transfer experiments, i.e., practicing the motor aspect of a task and transferring to a very different type of material. On transfer, *S*s in this group had to learn from scratch about the properties of our metric figures. Yet, on the second and third posttransfer trial there was no significant difference between the performance of this group and the other two experimental groups. By the third posttransfer trial the amount of transfer was approximately 80%, or only 7% less than the group having most transfer. It is felt that the comparisons between the posttransfer performance of Groups II and IV serve to place the effects of transfer among samples of the same class of material in their proper perspective. They show that although initially the differences between effects of this type of transfer and a more conventional one may be considerable, after little additional training the benefit of one need not be any greater than that of the other. It is worth noting at this time that all three experimental groups were still significantly different from the control group at Trial 6.

Lastly, it is now possible to assess the relative importance of different aspects of the total learning situation studied. How much, in other words, is the improvement of Group I between Trials 1 and 4 due to learning that is specific to the actual figures used, to the class of figures, and to the task, both in the perceptual and in the motor sense? Inasmuch as all three experimental groups were better on Trial 4 than the control group on Trial 1, but worse than the control group on Trial 4, the decrements and the savings may be used to estimate the relative importance of various components in learning. This amounts to an analytical sampling of the *state* of learning of the control group at a given trial and is not necessarily an examination of possibly concurrent learning processes.

The *S*s in the control group, in order to perform 1.11 sec. per figure better on Trial 4 than on Trial 1, must have learned motor and perceptual aspects of the situation, both of which in turn may be considered to have specific and general characteristics. Groups II and IV provided one line of evidence. The former were .66 sec. better on Trial 4 than the control on Trial 1, presumably by virtue of familiarity with the class of material plus both the perceptual and motor aspects of the identification task. The *S*s in Group IV were only .38 sec. better on Trial 4 than the control on Trial 1, presumably because all they had learned were the motor and perceptual aspects of the identification task. This leaves .28 sec. of the differences between control on Trial 1 and Group II on Trial 4 to be accounted for, but it is clear that part of the "saving" obtained by Group II can be attributed to learning of the task rather than of the class of material. What makes Group II worse on Trial 4 than the control group at that trial is lack of familiarity with the specific figures. This familiarity is, however, possessed by Group III, which in turn lacks nearly all the knowledge about the specific aspects of the identification task possessed by Group II. It is most interesting to find that in the statistical sense these two groups were not significantly different from each other throughout the transfer stage. It may be noted that Groups II and III complement each other in the sense that each has knowledge about two aspects of the situation which may in turn be called either specific or general, depending on whether one emphasizes the task or the material.

Meyer (7), in discussing the results of his learning-set studies, has suggested the terms "interproblem" and "intraproblem" learning. Although at first glance these terms are more operational than the terms specific and general, they do depend on an arbitrary definition of "problem." In the present task it seems reasonable to think in terms of two "problems": that which is largely perceptual and hinges around the ability to discriminate patterns, and that which is more predominantly motor, including both hand and eye movements, i.e., the ability to indicate the discrimination. For both these there can be clearly interproblem and intraproblem learning and for both there may well be the continuity of the process suggested by Meyer. The present data do not lend themselves to the same kind of quantification as that suggested by him, since the study involved a rather narrow scale in time and space, i.e., a few samples and training *Ss* for a few trials only. However, the material developed for the study of visual figure discrimination lends itself to extensive studies on human *Ss* comparable to those carried out with monkeys and prelanguage children by Harlow, Meyer, and others. It is more comparable to the material used in animal discrimination and reversal learning than were the nonsense syllables used in human learning-set studies by Meyer and Miles (8). They noted that *Ss* learning successively lists of nonsense syllables did not show as much benefit from progressive transfer as might have been expected on the basis of animal learning-set studies. It is possible that this was due to the increasing probability of intrusions from earlier lists having a detrimental effect on the learning of subsequent ones. In problem solving or identification tasks, where it may not be necessary to name a relation or an object, the probability of such intrusions having a detrimental effect is much less.

Summary

Four groups of 45 *Ss* each were used in an experiment to examine the effects of transfer between samples of the same class of material. It was shown that:

1. The *Ss* transferring between different samples of the same class of material (Group II) showed transfer effects which increased from 60% to 87% from the first to the third post-transfer trial when compared with a nontransferring control group (Group I).

2. Another group (IV) pretrained on the same task but using material of a different kind from that of the control group showed improvements in transfer from 34% to 80% from the first to the third transfer trial.

3. The performance of Groups II and IV differed significantly only on the first post-transfer trial.

4. Part of the transfer effect shown by Group II could be accounted for by familiarity with the task rather than the class of material used.

5. Pretraining on an identical sample of figures but on a different task (Group III) produced about the same transfer effects as pretraining on the same task but a different sample of figures (Group II).

References

1. Duncan, D. B. A significance test for differences between ranked treatments in an analysis of variance. *Va. J. Sci.*, 1951, **2**, 171–189.
2. Fitts, P. M., Weinstein, M., Rappaport, M., Anderson, N., & Leonard, J. A. Stimulus correlates of visual pattern recognition. *J. exp. Psychol.*, 1956, **51**, 1–11.
3. Gagné, R. M., Foster, H., & Crowley, M. E. The measurement of transfer of training. *Psychol. Bull.*, 1948, **45**, 97–130.
4. Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.
5. Harter, H. L., & Lum, M. D. Partially hierarchical models in the analysis of variance. USAF Wright Air Development Center, *Tech. Rep. WADC TR 55-33*, 1955.
6. McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* Longmans, Green, 1952.
7. Meyer, D. R. Food deprivation and discrimination reversal learning by monkeys. *J. exp. Psychol.*, 1951, **41**, 10–16.
8. Meyer, D. R., & Miles, R. C. Intralist-interlist relations in verbal learning. *J. exp. Psychol.*, 1953, **45**, 109–115.

(Received April 30, 1956)

Journal of Experimental Psychology
Vol. 53, No. 5, 1957

APPLICATION OF THE HALF-SPLIT TECHNIQUE TO PROBLEM-SOLVING TASKS [1,2]

ROBERT A. GOLDBECK,[3] BENJAMIN B. BERNSTEIN,[4]
W. A. HILLIX, AND MELVIN H. MARX

University of Missouri

The need to maintain complex electronic equipment has stimulated interest in a class of problem solving called trouble shooting. Faulty components in electronic systems must be located on the basis of the performance of the system and its parts. Solution of a problem may be considered to have two parts: the determination of a set of possible faulty components, and the elimination of all but the component that is actually faulty through some checking procedure.

Miller, Folley, and Smith (3) have presented a logical demonstration of the superior efficiency of the half-split technique for problems so considered. Each check eliminates half of the alternatives. However, there has been no empirical examination of this solution method, the effectiveness of which will depend upon human capability to use it.

First, a set of trouble alternatives (possible faulty components) must be determined. Then check points which divide the set of alternatives into halves must be located. If the trouble alternatives are directly given and have a simple structure which can be easily divided into halves, empirical results should follow closely the mathematics of information theory. But if either of these two aspects of the method is complicated by complexities in the structure of the system, method effectiveness should be more intimately related to the capabilities of *Ss*.

The two experiments reported here were designed to provide empirical tests of application of the half-split method. The major purpose of our first experiment was to examine the effects of problem complexity upon the efficiency of solution for *Ss* instructed in the half-split method and for *Ss* not so instructed. Specifically, the experiment was designed (*a*) to test the hypothesis that half-split instruction increased efficiency for simple problems, and (*b*) to determine half-split efficiency for more complex problems. As implied in the preceding paragraph, it was expected that the half-split instructions would improve efficiency for the simpler problems, but not necessarily for the more complex.

The second experiment was designed to determine whether additional practice, different methods of

[1] This research was supported in part by the United States Air Force under Contract No. AF 18(600)-1053 monitored by the Maintenance Laboratory, Air Force Personnel and Training Research Center, Lowry Air Force Base, Denver, Colorado. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government.

[2] Experiment I represents the major part of a doctoral dissertation by the first-named author at the University of Missouri. The second-named author was primarily responsible for Exp. II. The third-named author is primarily responsible for the form of the present report. The fourth-named author served as principal investigator and supervisor of the research.

[3] Now at American Institute for Research, Pittsburgh, Pennsylvania.

[4] Now at Bombing-Navigation Research Unit, Operator Laboratory Field Unit No. 4, Air Force Personnel and Training Research Center, Mather Air Force Base, California.

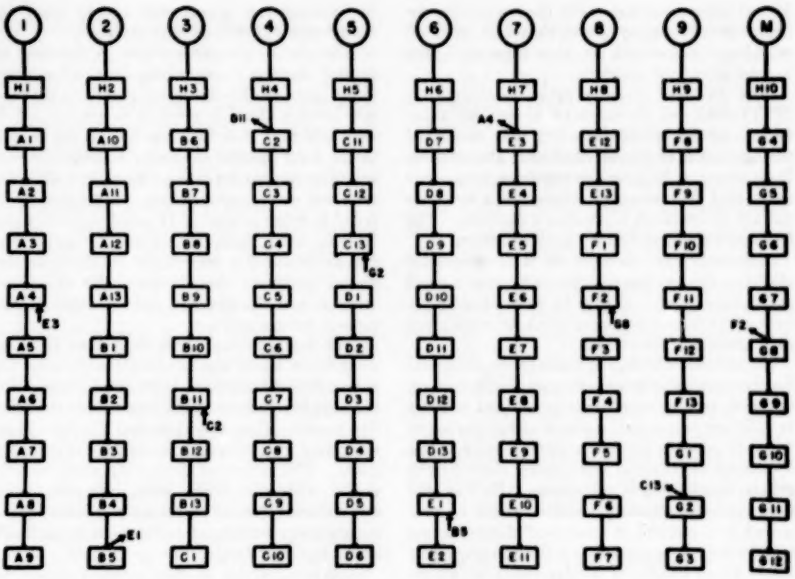


FIG. 1. Complexity I diagram.

instruction, and more complex training materials would produce an improvement in *S*'s ability to arrive at a set of alternatives, and, if so, whether the half-split method would then be more efficient. Four different training methods were tested for effectiveness.

EXPERIMENT I

Method

Apparatus.—The problems were presented to *S* by means of an apparatus designed to simulate an electronic system, and a diagram of the system relationships. The system consisted of 10 display items and 100 connected components. The apparatus is described elsewhere (2) in more detail.

The diagram of the simplest system is shown in Fig. 1. Each of the 10 display items terminated a signal chain consisting of 10 units connected in series. Each signal chain was connected to one of the other signal chains at a random point in the chain. Thus each simple chain had one remote connection; this was either an input to the chain or an output from it. This system diagram, used in Exp. I at Complexity I, formed the skeleton for the other two diagrams. For the Complexity II diagram, an additional remote connection was made for each of the chains. For Complexity III, a third connection was made. As complexity increased, the signal network feeding the display items became more complicated and a faulty unit caused more symptoms (display items that were off) in the symptom array.

The display items were presented as nine lights and a meter on a panel. When operating properly, the components transmitted a signal to the display items connected to them, causing the nine lights to be on and the meter to give a specified reading.

A second panel consisted of switches and an associated light. By throwing the switch representing the check point of any one of the 100 components under the display items, *S* could find out whether or not the signal was present at the output of the component. If it was, the light would come on; if not, the light would remain off.

There was a third panel which also had a switch for each of the 100 components. When *S* thought he knew the faulty component, he could throw the associated switch on this third panel, thus effectively replacing this component.

If that component was really the faulty one, the items on the display panel that had been off would now come back on, thus showing *S* that he had solved the problem.

The *Ss* were given a problem-solving test (PST) designed to measure ability on those aspects of the trouble shooting that concerned the deduction of the set of trouble alternatives. It consisted of 20 questions requiring *S* to count units that had specified relationships to other units in the network printed on a diagram. The 64 units were identified by pairs of letters.

Subjects.—The *Ss* were 90 male university students drawn from a general experimental psychology class. Thirty *Ss* were assigned to each of the three complexity levels; 15 were given each method of instruction.

Procedure.—On Day 1, *S* was given a diagram for the complexity level to be used during testing, but with remote connection inputs and outputs at different points in the simple signal chains, in order to prevent specific transfer effects. The *S* listened to recorded instructions covering the general functioning of the system. He was told that a trouble-shooting problem would be presented as a pattern of on-and-off display items called the symptom array. A faulty component would cause whichever display items were connected to it to be off. By tracing the relevant connections on the system diagram, *S* could determine which units might cause the given pattern of symptom items to be off. He could also determine the most efficient place to check. He could check points in the system by means of switches on the checking panel. Finally, the checking information would enable him to locate the trouble source. The problem was solved when he threw the replacement switch for the faulty component, and the display items came back on. After hearing the instructions, *S* was allowed to ask questions and then required to answer a list of questions to check his comprehension of the task. These questions required *S* to perform and interpret correctly specified checks and replacements.

On Day 2, *Ss* were given a second set of check-out questions to insure retention of the instructions. Then *Ss* in the half-split (HS) groups listened to a 5.5-min. recording of instructions on the half-split method and were asked if they had any questions about the method. Check-out questions followed which required *S* to correctly apply the half-split method, step by step, to two simple problems. The *S* was to examine the symptom array and the associated interconnected units. For any given array, there were several units whose malfunctioning could produce that array. These were called trouble-possible units. The *S* was to locate these and successively halve them by making the appropriate checks until the faulty unit was isolated and replaced.

The *Ss* in the control groups received no special method instructions, but were given those parts of the check-out questions that did not involve the HS method as such. All *Ss* were told that their task was to find the trouble in the least number of checks possible; 20 min. would be allowed for each problem, but otherwise time was not a consideration. A total of 12 *Ss* failed to finish a total of 17 problems within the 20-min. time limit. Estimates of times and checks to solution were made for these incompleted problems, based upon the degree of solution and the average rate at which *S* progressed toward solution.

The first problem given each *S* at all three complexity levels was a trouble with only one symptom—the simplest type of problem. For the next five problems all *Ss* were given the same five troubles (i.e., the same five "faulty units" were used regardless of method or complexity level). The symptom array for a given trouble varied with complexity level. Trouble order was counterbalanced according to a latin square arrangement with three replications in each cell of the factorial design.

Following testing on the electronic system *Ss* were given the PST.

Results

Checking scores and times.—The over-all method difference was not significant ($F < 1$) for a simple analysis of variance using method, complexity, and within-*Ss* variance. The Method × Complexity interaction was also not significant ($F < 1$). As was expected, the complexity difference was significant ($F = 3.89$, $P < .05$ for $df = 2, 84$); the number of checks was positively related to complexity.

Since a difference in favor of the HS group had been predicted at Complexity I, a separate analysis was carried out for Complexity I only. Here the method difference was significant ($F = 4.68$, $P < .05$ for $df = 1, 24$).

The solution time for the HS groups was significantly longer than for the control ($F = 7.44$, $P < .01$ for $df = 1, 84$). This reflects the addi-

tional covert solution behavior associated with the attempt to reduce checking responses by means of the half-split procedure. With increasing complexity the times increased ($F = 18.60$, $P < .001$ for $df = 2, 84$), and the obtained HS-control difference became more marked.

PST analysis.—In order to assess the effects of levels of PST performance in conjunction with the main variables, *Ss* were divided into three levels according to the upper, middle, and lower third of the PST score distribution for each group.[5] An analysis of variance was performed using these levels as an additional variable. The results agreed with those of the previous analysis and produced two additional findings. PST level had a highly significant effect ($F = 9.20$, $P < .001$ for $df = 2, 72$), as did the PST Level × Complexity interaction ($F = 3.49$, $P < .05$ for $df = 4, 72$). Increasing complexity resulted in increased check scores only for *Ss* scoring low on the PST.

It would seem that the difficulty of complex problems was in large part dependent upon *S* characteristics which were measured by the PST.

The product-moment correlations between PST and checking scores at the three complexity levels were as follows: for the half-split group, + .16, − .31, and − .80; for the control group, + .24, − .63, and − .80. The increase in correlation with complexity indicates a greater role for PST ability on complex problems. High PST scores and low checking scores represent good performance, producing negative *r*'s.

Since the PST was administered to *Ss* in Exp. I after they had performed the apparatus problems, it was possible that differential transfer from the apparatus problems to the PST occurred for the different experimental groups. To test this possibility, an analysis of variance was carried out on the PST scores classified according to the complexity and method variables used for the apparatus problems. There was no evidence for differential transfer ($F < 1$ for M and C, 1.48 for M × C with $df = 2, 84$ for the latter).

Reliability.—The reliability of the total-checks scores for each of the major experimental groups was determined by an analysis of variance technique developed by Hoyt (1). For the HS groups the reliability values were .62, .72, and .53 for Complexities I, II, and III, respectively. For the control groups the reliability values were .50, .99, and .74. These values indicate reasonably high reliability for the experimental situation. The split-half reliability using an odd-even breakdown on the PST was .83.

Half-split performance measures.—The *Ss* were unable to benefit from instruction in the logically superior HS method, at least at the higher complexities. It is then appropriate to examine the reasons for this failure and to determine which aspects of the method present difficulty.

Since HS *Ss* were instructed first to determine the units which if faulty could cause the symptom array presented (trouble-possible units), the number of times each *S*'s first check

[5] The levels breakdown of PST scores by groups, some analysis of variance tables mentioned in the text, the instructions for both experiments, and certain tabulations of data from both experiments, have been sent to the American Documentation Institute. Order Document No. 5145 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $2.00 for microfilm or $3.75 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

was among the trouble-possible units was scored. The HS groups showed a slight advantage in beginning their responses in the trouble-possible area, but the methods effect was not significant ($F = 2.29$ for $df = 1, 72$). It is surprising that *Ss* who have been told how to go about limiting the scope of trouble alternatives were not significantly better at this part of the task. A strong effect of PST levels ($F = 5.66$, $P < .01$ for $df = 2, 72$) indicates that there was an important ability factor involved in determining trouble-possible units.

A significant complexity by PST level interaction ($F = 2.50$, $P < .05$ for $df = 4, 72$) resulted from a steeper decline in trouble-possible first checks with increasing complexity for low PST-scoring *Ss* than for middle- or high-scoring *Ss*. However, for the high PST level the complexity difference was still significant ($F = 4.24$, $P < .05$ for $df = 2, 20$). These results, plus the significant complexity effect ($F = 10.04$, $P < .001$ for $df = 2, 72$), indicate that increasing complexity had the effect of preventing *Ss* from accomplishing the first step in the half-split method used in this experiment. At the higher complexities *Ss* were limited by an ability factor in arriving at the trouble-possible set of alternatives by an interpretation of the symptom array.

The second part of the HS method concerned making half-split checks among the trouble-possible units. Scoring each *S* for the number of perfect half-split checking sequences used in the trouble-possible area resulted in the superiority of the HS groups. The method difference was highly significant ($F = 30.65$, $P < .001$ for $df = 1, 72$). Here there was no significant complexity difference. However, with increasing complexity the number of trouble-possible units was generally smaller, requiring fewer steps in the half-split checking sequence.

To eliminate this bias, only the first check in the trouble-possible area was considered. An analysis of variance using this measure showed a significant method effect ($F = 71.12$, $P < .001$ for $df = 1, 72$) and a Complexity by Method interaction ($F = 4.57$, $P < .05$ for $df = 2, 72$). This resulted from a reduction, in half-split checks, with increased complexity by the HS group. Since the trouble-possible units were contained in simple lines at all complexities (except for two troubles at Complexity II where two lines were involved), this reduction in half-split checks can be attributed to the difficulty in the first step of getting a set of alternatives. If the trouble alternatives cannot be recognized as such, they cannot be split in half.

Correct solutions.—Although primary emphasis has been placed upon the number of checks used to solve a problem, consideration of whether or not *Ss*' first solution was a correct one is of some importance. When each *S* was scored for the number of correct first replacements for the five problems, only PST level showed a significant effect ($F = 5.99$, $P < .01$ for $df = 2, 72$). This indicates that this aspect of performance was determined by the characteristics of *S* rather than by the situational variables.

In addition, each *S* was scored for the average number of checks used for finding troubles where he made no incorrect replacements. There was a significant method effect ($F = 23.47$, $P < .001$ for $df = 1, 72$) and significant Method × Complexity interaction ($F = 5.77$, $P < .01$ for $df = 2, 72$). The HS group was most superior at Com-

plexity III. Where *S* avoided any incorrect solutions, the half-split method was more effective.

The results of Exp. I may be summed up as follows: (*a*) The increase in number of checks required with increased complexity indicates the importance of being able to deduce the trouble-possible units correctly, especially since the number of trouble-possible units decreased with increasing complexity. If *S*s could have conceptualized system relationships perfectly, fewer checks would have been required. (*b*) The HS method cannot in general be said to be superior. (*c*) There is evidence that it is superior under certain conditions, such as with diagrams of low complexity, where *S*s apparently had little difficulty in finding the trouble-possible units, and where *S*s avoided incorrect replacements.

This latter finding gave some hope that if all *S*s could be given increased training on working with the system relationships, then the HS method would be superior even with more complex diagrams. Experiment II was an attempt to improve the effectiveness of the HS method by improving instruction and giving extended practice.

Experiment II

Method

Apparatus.—The same apparatus was used as in Exp. I, except that only the system diagram of greatest complexity (three remote connections) was used during test trials. The training diagram was different, being the previous test diagram of intermediate complexity (two remote connections).

Subjects.—The *S*s were 40 male students from a general experimental psychology class at the University of Missouri. Ten *S*s were randomly assigned to each of the four instructional groups.

Procedure.—Before beginning either the instructions or the experimental sessions in trouble-shooting, all *S*s were given the PST for use in analysis of covariance. The basic instructional procedure was similar to that of Exp. I. However, all *S*s had 2-hr. sessions on the first day, rather than the 1-hr. session of Exp. I. All *S*s had additional practice with the diagram of intermediate complexity, with special conditions of practice for each group.

Group C (control) was essentially like the control group of Exp. I, except for the additional practice. During this practice, *E* asked *S* to specify the check he wanted made. If it did not violate any basic instructions, *S* was allowed to make the check.

Group DHS (Deductive Half-Split) was like the previous Group HS, with additional practice on the more complex diagrams. During the supervised practice for this group, *S* was allowed to make the check only if he had correctly deduced the trouble-possible items, and if the check correctly half-split those units. If it was an incorrect check, *S* was told only that it was incorrect because it was not a half-split check, whereupon *S* kept trying until the correct check was made.

Group DA (Deductive Alternatives) differed from the groups previously used. The rationale for setting up this new group is as follows: It is not necessary to arrive at the *minimum* set of trouble-possible units in order to find the trouble, so long as the set arrived at contains the faulty unit. It may be that *S* can do better where he is required only to find *some* set of alternatives that includes the trouble spot, if finding the minimum set would strain his capabilities, as happened in Exp. I. Thus, Group DA was instructed in three alternative ways of conceiving of the set of alternatives, and allowed to choose any one of them for half-split application during the test. The minimum set consisted *only* of the trouble-possible components, just as in Group DHS. The maximum set consisted of all components in the lines beneath display items that were off. The intermediate set was obtained by eliminating those lines in the maximum set which did not have components feeding into other off-display items. For example, if there were three display items off, all the components beneath the off-display items (30 components) would constitute the maximum set. If one of the three lines involved did not have components feeding into either of the other two off-display items, it could not possibly contain the faulty component and thus could be eliminated from consideration for the intermediate set, which would then consist of 20 components. Further considerations might allow elimination of further components for the determination of the minimum set. During the practice session for this group, *E* told *S* what kind of set to split and allowed only a half-split check of the designated set.

Group HS (Half-Split) received varied practice in half-splitting sets of components, but was given no practice in determining the sets of alternatives. The symptom array was not used during practice for this group. The *E* told *S*

TABLE 1

GROUP MEANS FOR PERFORMANCE AND PST MEASURES, EXP. II

Measure	Group			
	C	DHS	DA	HS
Mean checks to solution	57.1	22.3	39.1	60.8
Mean trouble-possible first checks	1.4	3.8	1.8	0.1
Mean PST scores	11.4	14.1	12.4	8.9

which units to split, and corrected him only if checks did not correctly half-split the given set of alternatives.

Test procedure was like that in Exp. I. There were two trouble-by-trial latin square arrangements in each group. The *S*s were allowed 25 min. to finish a problem.

Results

Checking scores and times.—Both the mean number of checks required to find the five troubles and the mean number of times *S*'s first check was among the trouble-possible units are shown in Table 1, along with the group means on the PST. The groups scored in the same order on all three measures, and the group difference on the PST was almost significant ($F = 2.62$, with 2.86 required for $P = .05$, $df = 3, 36$), despite the fact that *S*s were randomly assigned.

Therefore, an analysis of covariance was used to determine whether the groups differed due to treatments, in addition to the variation correlated with the PST. Since the means and variances for number of checks to solution were correlated, and the variances were significantly heterogeneous ($\chi^2 = 10.99$, $P < .02$), a log transformation was used and successfully corrected for the heterogeneity. The method difference for the transformed scores was significant ($F = 6.75$, $P < .01$ for $df = 3, 35$). The adjusted log means for the groups were as follows: C, 1.68; DHS, 1.39; DA, 1.59; and HS, 1.66. Tests of these means, using a generalized error term of .069 derived from the analysis of covariance, showed that Group DHS was superior to each of the other groups at the .01 level. None of the other comparisons was significant.

The analysis of covariance of the trouble-possible first check scores yielded a highly significant treatment difference ($F = 13.22$, $P < .001$ for $df = 3, 36$). The adjusted means for the four groups were: C, 1.44; DHS, 3.50; DA, 1.71; and HS, .46. The *t* tests showed the DHS group superior to all other groups at beyond the .001 level. The HS group was inferior to all other groups at beyond the .05 level. The only insignificant difference was between Groups DA and C. This result indicates that the different instructions and conditions of practice differed in their effectiveness in helping *S* to accomplish the first step of the trouble-shooting process. The difference in number of checks to solution seems attributable to differences in ability to deduce a set of trouble-possible units, rather than to differences in ability to half-split alternatives, once determined. The analysis of covariance on time to solution did not yield any significant results.

Another subsidiary measure of efficiency was the number of times *S* "replaced" a unit on the replacement panel. Perfect efficiency would involve using only as many replacements as there were troubles, five in this case. The average numbers were: C, 9.0; DHS, 8.0; DA, 10.1; and HS, 12.1. The analysis of covariance of these scores did not approach significance.

PST scores.—Product-moment correlations between number of checks to solution and PST scores for the four groups were as follows: C, − .46;

DHS, − .59; DA, − .56; and HS, − .59. The required r for significance at the .05 level is .60 for the small N of 10. The correlations were, of course, negative since low check scores indicate better performance than high check scores.

Comparison with Exp. I.—Groups C and DHS from Exp. II were combined with the two groups from Exp. I in an analysis of covariance with PST scores as the predictor variable. This use seems justified despite the fact that PST was administered *after* the apparatus test in Exp. I, and *before* the apparatus test in Exp. II, since there was no evidence of differential transfer in Exp. I.

The analysis showed a significant Study×Method interaction ($F=6.34$, $P < .02$ for $df = 1, 45$). Since the variances were again heterogeneous ($\chi^2 = 12.85$, $P < .01$), a conservative interpretation is necessary. Even so, this comparison tends to support the observation that the instructions and practice given Group DHS in Exp. II were more effective in reducing the number of checks than were the less extensive instructions in Exp. I. The mean number of checks for Exp. I at the high complexity were C, 47.8; HS, 45.3; for Exp. II, C, 57.1; DHS, 22.3.

The analysis of covariance of the trouble-possible first check scores for the two studies also showed a significant Study × Method interaction ($F = 6.09$, $P < .02$ for $df = 1, 45$). Here the variances were not heterogeneous ($\chi^2 = 5.30$, $P > .10$). The method of instruction seemed to produce a dramatic change in the ability to deduce a set of trouble-possible units as well as in the ability to locate a faulty unit in a minimal number of checks.

Discussion

Although the half-split method has been demonstrated to be logically superior (3), the results of the two experiments herein reported indicate that its empirical superiority is a function of the nature of the task, and the degree and type of training given the trouble shooter. The *S*'s checks more closely approximate those which would be expected logically when task complexity is low or instruction in deducing trouble-possible units is high.

There seem to be two components in the empirical process of trouble shooting. The first component involves finding the trouble-possible units. The second component involves half-splitting the trouble-possible units at each stage by some checking operation, or at least eliminating part of the units from consideration. In the present experiments the second step was relatively simple. Those *S*s who were able to locate the trouble-possible units, as measured by the occurrence of the first check in the trouble area, were able to perform efficiently in finding the troubles.

Wherever a significant difference in number of checks was found between groups, there was an accompanying difference between the number of initial checks in the trouble-possible area to help explain it. In Exp. I, failures to find expected differences were correlated with difficulty in isolating trouble-possible units. Also, half-split *S*s who avoided making any incorrect replacements were more efficient than comparable control *S*s. The importance of the deductive component was further substantiated by the marked superiority of DHS *S*s in Exp. II as compared to HS *S*s. This difference seems attributable to their greater practice in the deductive component.

It appears then that the failure to find a general superiority for the half-split method is the result of *S*'s inability to perform efficiently in the more fundamental deductive task of isolating sets of trouble units. This result suggests that it will be more profitable in future research on training methods to emphasize the more basic deductive phase rather than the relatively mechanical half-splitting aspect of the process.

Summary

This is a report of two experiments on the use of the half-split method of locating trouble sources in malfunctioning equipment. The experiments were carried out using apparatus and diagrams designed especially for this purpose. Experiment I investigated the effects of differing complexity of system and differing instructions on efficiency in locating trouble sources. Experiment II was a further investigation of the effects of instruction and practice for the high complexity condition.

Both sets of results indicated that where the system relationships were easily mastered by a given *S* the half-split method was an aid to efficiency. However, *S* needed either relatively high ability or instructional aid to overcome the load put on his capacities by the more complex systems. He had to be able to deduce from the symptom complex which units might be faulty in order to use the half-split method efficiently. It is probable that this ability also indicated that these *S*s were better able to understand the meanings of the checks made. It was concluded that deductive ability is a prerequisite for application of the half-split technique, and indeed may play a preponderant role at each stage of the trouble-shooting task.

References

1. Hoyt, C. J. Test reliability estimated by analysis of variance. *Psychometrika*, 1941, **6**, 153–160.
2. Marx, M. H., Goldbeck, R. A., & Bernstein, B. B. An apparatus for investigating the methods humans use in solving complex problems. *Amer. J. Psychol.*, 1956, **69**, 462–465.
3. Miller, R. B., Folley, J. P., Jr., & Smith, P. R. Systematic trouble shooting and the half-split technique. Lackland Air Force Base, Texas: Human Resources Research Center, July, 1953. (*Tech. Rep.* 53-21).

(Received May 4, 1956)

Journal of Experimental Psychology
Vol. 53, No. 5, 1957

THE RELATION OF EYE MOVEMENTS DURING SLEEP TO DREAM ACTIVITY: AN OBJECTIVE METHOD FOR THE STUDY OF DREAMING

WILLIAM DEMENT[1] AND NATHANIEL KLEITMAN[2]

Department of Physiology, University of Chicago

The study of dream activity and its relation to physiological variables during sleep necessitates a reliable method of determining with precision when dreaming occurs. This knowledge, in the final analysis, always depends upon the subjective report of the dreamer, but becomes relatively objective if such reports can be significantly related to some physiological phenomena which in turn can be measured by physical techniques.

Such a relationship was reported by Aserinsky and Kleitman (**1**) who observed periods of rapid, conjugate eye movements during sleep and found a high incidence of dream recall in *S*s awakened during these periods and a low incidence when awakened at other times. The occurrence of these characteristic eye movements and their relation to dreaming were confirmed in both normal *S*s and schizophrenics (**4**), and they were shown to appear at regular intervals in relation to a cyclic change in the depth of sleep during the night as measured by the EEG (**5**).

This paper represents the results of a rigorous testing of the relation between eye movements and dreaming. Three approaches were used: (*a*) Dream recall during rapid eye movement or quiescent periods was elicited without direct contact between *E* and *S*, thus eliminating the possibility of unintentional cuing by *E*. (*b*) The subjective estimate of the duration of dreams was compared with the length of eye movement periods before awakening, reasoning that there should be a positive correlation if dreaming and eye movements were concurrent. (*c*) The pattern of the eye movements was related to the dream content to test whether they represented a specific expression of the visual experience of dreaming or merely a random motor discharge of a more active central nervous system.

[1] Postdoctoral Public Health Service Research Fellow of the National Institute of Neurological Diseases and Blindness.

[2] Aided by a grant from the Wallace C. and Clara A. Abbott Memorial Fund of the University of Chicago.

METHOD

The *S*s for the experiments were seven adult males and two adult females. Five were studied intensively while the data gathered from the other four were minimal with the main intent of confirming the results on the first five.

In a typical experiment, *S* reported to the laboratory a little before his usual bedtime. He was instructed to eat normally but to abstain from alcoholic or caffeine-containing beverages on the day of the experiment. Two or more electrodes were attached near the eyes for registering changes in the corneoretinal potential fields as the eyes moved. Two or three electrodes were affixed to the scalp for recording brain waves as a criterion of depth of sleep. The *S* then went to bed in a quiet, dark room. All electrode lead wires were further attached to the top of the head and from there to the lead box at the head of the bed in a single cord to minimize the possibility of entanglement and allow *S* a free range of movement. The potentials were amplified by a Model III Grass Electroencephalograph in an adjoining room. The electroencephalograph was run continuously throughout the sleep period at a paper speed of 3 or 6 mm. per sec. which allowed easy recognition of eye-movement potentials. A faster speed (3 cm./sec.) was used for detailed examination of the brain waves although the slower speed permitted at least an approximate

TABLE 1

SUMMARY OF EXPERIMENTS

Ss	Nights Slept	Awakenings	Average Nightly Awakenings	Average Sleeping Time
DN	6	50	8.3	7:50
IR	12	65	5.4	4:20
KC	17	74	4.4	6:00
WD	11	77	7.0	6:30
PM	9	55	6.1	6:20
KK	2	10	5.0	6:00
SM	1	6	6.0	6:40
DM	1	4	4.0	7:00
MG	2	10	5.0	6:10
Totals	61	351	5.7	6:00

estimation of the gross pattern. The criteria of eye-movement potentials and their differentiation from brain wave artifacts have been discussed at length elsewhere (**1, 4**).

At various times during the night *Ss* were awakened to test their dream recall. The return to sleep after such an awakening invariably took less than 5 min. Table 1 is a summary of the experiments showing the number of nights each *S* slept and the number of awakenings. In all, 21% of the awakenings fell in the first 2 hr. of sleep, 29% in the second two, 28% in the third two, and 22% in the fourth two.

RESULTS

The occurrence of rapid eye movements.[3]—Discrete periods during which their eyes exhibited rapid movements were observed in all nine *Ss* every night they slept. These periods were characterized by a low-voltage, relatively fast pattern in the EEG. The interspersed periods in which rapid eye movements were absent showed EEG patterns indicative of deeper sleep, either a predominance of high-voltage, slow activity, or frequent, well-defined sleep spindles with a low-voltage background. No REM's were ever observed during the initial onset of sleep although the EEG always passed through a stage similar to that accompanying the rapid eye movement periods occurring later in the night. These findings concerning associated EEG patterns were identical with previous observations on uninterrupted sleep (**5**).

An accurate appraisal of the mean duration of the REM periods was impossible since most were terminated artificially by an awakening. However, those that were not so terminated varied between 3 and 50 min. in duration with a mean of about 20 min., and they tended to be longer the later in the night they occurred. The eyes were not constantly in motion during such periods; rather, the activity occurred in bursts of one or two, up to fifty or a hundred movements. A single movement was generally accomplished in .1–.2 sec. and was followed by a fixational pause of varying duration. The amount, pattern, and size of the movements varied irregularly from period to period.

The REM periods occurred at fairly regular intervals throughout the night. The frequency of occurrence seemed to be relatively constant and characteristic for the individual. DM and WD averaged one eye-movement period every 70 min. and every 75 min. respectively. KC averaged one eye-movement period every 104 min. The other *Ss* fell between these two extremes. The average for the whole group was one REM period every 92 min.

Despite the considerable disturbance of being awakened a number of times, the frequency and regularity with which REM periods occurred was almost exactly comparable to that seen previously in a study of uninterrupted sleep (**5**). If the awakening occurred during a NREM period, the return to sleep was never

[3] In most of the remaining text the following abbreviations will be used: REM's (rapid eye movements) and NREM's (no rapid eye movements).

associated with REM's, nor was the time of onset of the next REM period markedly changed from that which would have been expected in the absence of an awakening. An awakening during an REM period generally terminated the REM's until the next period, and the sequence of EEG changes, excluding the brief period of wakefulness, was the same as that following an REM period that ended spontaneously. Exceptions occurred when *S* was awakened during an REM period in the final hours of sleep when the period was likely to be quite long if uninterrupted. On these occasions, the REM's sometimes started up again when *S* fell asleep. It seemed as though a period of heightened CNS activity had not run its normal course and, although *S* was able to fall asleep, he continued to dream.

Eye movement periods and dream recall.—For all awakenings to elicit dream recall, the arousing stimulus was the ringing of an ordinary doorbell placed near the bed and sufficiently loud to ensure immediate awakening in all levels of sleep. The *S*s then spoke into a recording device near the bed. They were instructed to first state whether or not they had been dreaming and then, if they could, to relate the content of the dream. When *S* had finished speaking *E*, who could hear their voices, occasionally entered the room to further question them on some particular point of the dream. There was no communication between *S* and *E* in any instance, it must be emphasized, until *S* had definitely commited himself. The *S*s were considered to have been dreaming only if they could relate a coherent, fairly detailed description of dream content. Assertions that they had dreamed without recall of content, or vague, fragmentary impressions of content, were considered negative.

TABLE 2

INSTANCES OF DREAM RECALL AFTER AWAKENINGS DURING PERIODS OF RAPID EYE MOVEMENTS OR PERIODS OF NO RAPID EYE MOVEMENTS

S	Rapid Eye Movements		No Rapid Eye Movements	
	Dream Recall	No Recall	Dream Recall	No Recall
DN	17	9	3	21
IR	26	8	2	29
KC	36	4	3	31
WD	37	5	1	34
PM	24	6	2	23
KK	4	1	0	5
SM	2	2	0	2
DM	2	1	0	1
MG	4	3	0	3
Totals	152	39	11	149

The awakenings were done either during REM periods or at varying increments of time after the cessation of eye movements during the interspersed periods of NREM's. The *S*s, of course, were never informed when awakened whether or not their eyes had been moving.

Table 2 shows the results of the attempts to recall dreams after the various awakenings. The REM or NREM awakenings for PM and KC were chosen according to a table of random numbers to eliminate any possibility of an unintentional pattern. For DN, a pattern was followed: first three REM awakenings, then three NREM awakenings, and so on. WD was told he would be awakened *only* when the recording indicated that he was dreaming, but REM and NREM awakenings were then interspersed randomly The type of awakenings for IR was chosen according to the whim of *E*

The *S*s uniformly showed a high incidence of dream recall following

REM awakenings and a very low incidence of recall following awakenings during periods of NREM's regardless of how the awakenings were chosen. In particular, DN was not more accurate than the others although there was a pattern he might have learned, and WD was not less accurate although he was deliberately misled to expect to have been dreaming every time he was awakened. Over a narrow range, some *S*s appeared better able to recall dreams than others.

Table 3 compares the results of the first half of the series of REM awakenings with the last half. Practice was certainly not a significant factor as only one *S* showed any degree of improvement of recall on later nights as compared with the early ones.

The incidence of dream recall dropped precipitously almost immediately upon cessation of REM's. In 17 NREM awakenings that were done within 8 min. after the end of a REM period, 5 dreams were recalled. Although small, this was a much higher incidence of dream recall than occurred when the NREM awakenings followed the end of REM periods by *more* than 8 min. In the latter category only 6 dreams were recalled in 132 awakenings.

TABLE 3

COMPARISON OF FIRST HALF OF SERIES OF RAPID EYE MOVEMENT AWAKENINGS WITH SECOND HALF

S	First Half		Second Half	
	Dream Recall	No Recall	Dream Recall	No Recall
DN	12	1	5	8
IR	12	5	14	3
KC	18	2	18	2
WD	19	2	18	3
PM	12	3	12	3
Total	73	13	67	19

In general, *S*s were best able to make an emphatic statement that they had not been dreaming when the NREM awakenings were done during an intermediate stage of sleep as indicated by a brain-wave pattern of spindling with a low-voltage background. When aroused during a deep stage of sleep characterized by high-voltage, slow waves in the EEG, *S*s often awoke somewhat bewildered. In this state they frequently felt that they must have been dreaming although they could not remember the dream or, on the other hand, that they had not been asleep at all. They sometimes had a great variety of feelings to describe—such as pleasantness, anxiety, detachment, etc., but these could not be related to any specific dream content.

Most of the instances of inability to recall dreaming after awakenings during REM periods occurred in the early part of the night. Of 39 negative reports in the entire study, 19 occurred after awakenings during REM periods falling in the first 2 hr. of sleep, 11 after REM awakenings during the second 2 hr., 5 in the third 2 hr., and 4 in the last 2 hr. There was no such variation relating to awakenings during the interspersed periods of ocular quiescence, the incidence of dream recall being uniformly low, regardless of whether the early or late part of the night was being considered.

Length of rapid eye movement periods and subjective dream-duration estimates.—If the length of the REM periods were proportional to the subjectively estimated duration of the dreams, it would further help to establish the relatedness of the two and would give some information about the rate at which dreaming progresses.

At first, *S*s were awakened at

TABLE 4

RESULTS OF DREAM-DURATION ESTIMATES AFTER 5 OR 15 MIN. OF RAPID EYE MOVEMENTS

S	5 Minutes		15 Minutes	
	Right	Wrong	Right	Wrong
DN	8	2	5	5
IR	11	1	7	3
KC	7	0	12	1
WD	13	1	15	1
PM	6	2	8	3
Total	45	6	47	13

various increments of time after the REM's had begun and were requested to estimate to the nearest minute the amount of time they had been dreaming. This proved to be too difficult, although the estimates were always of the same order of magnitude as the lengths of the REM periods, and were occasionally exactly right.

A series was then done in which *Ss* were awakened either 5 or 15 min. after the onset of REM's and were required on the basis of their recall of the dream to decide which was the correct duration. The 5- or 15-min. periods were chosen on the basis of a random series. Table 4 shows the results of these awakenings. All *Ss* were able to choose the correct dream duration with high accuracy except DN. This *S*, however, made most of his incorrect choices by estimating 15 min. to be 5 min. This is consistent with the interpretation that the dream was longer, but he was only able to recall the latter fraction and thus thought it was shorter than it actually was.

In addition to depending on the amount of actual dreaming, the lengths of the dream narratives were undoubtedly influenced by many other factors as, for example, the loquacity or taciturnity of *S*. However, the lengths of the dream narratives still showed a significant relationship to the duration of REM periods before awakening. Table 5 shows the correlations between minutes of REM's and lengths of dream narratives for each *S*. The number of words in the narrative was the measurement of length. Of the 152 dreams recalled, 26 were not included because poor recording did not allow complete transcription. Dream narratives recalled after 30 or as much as 50 min. of REM's were not a great deal longer than those after 15 min. although *Ss* had the impression that they had been dreaming for an unusually long time. This was perhaps due to inability to remember all the details of very long dreams.

Specific eye-movement patterns and visual imagery of the dream.—The quality and quantity of the REM's themselves showed endless variation. There was much or little movement, big or small movements, and so on. As has been stated, the movements occurred in bursts of activity separated by periods of relative inactivity. However, the brain-wave stage during the whole period remained the same whether there was much or little movement at any given moment of the period.

It was hypothesized that the movements represented the visual imagery of the dream, that is, that they

TABLE 5

CORRELATION BETWEEN DURATION OF REM PERIODS IN MINUTES AND NUMBER OF WORDS IN DREAM NARRATIVES

Subjects	Number of Dreams	*r*	*P*
DN	15	.60	< .02
IR	25	.68	< .001
KC	31	.40	< .05
WD	35	.71	< .001
PM	20	.53	< .02

corresponded to where and at what the dreamer was looking. An attempt to account for every movement by having *S* state chronologically in what directions he had gazed in the dream proved futile. The *S*s could not recall the dream with such a high order of detail and precision.

In a slightly different approach, *S*s were awakened as soon as one of four predominant patterns of movement had persisted for at least 1 min. and were asked to describe in detail the dream content just before awakening. The four patterns were: (*a*) mainly vertical eye movements, (*b*) mainly horizontal movements, (*c*) both vertical and horizontal movements, and (*d*) very little or no movement. The prevalence of the horizontal or vertical components was determined by placing leads both vertically and horizontally around the eyes.

A total of 35 awakenings was accumulated from the nine *S*s. Periods of either pure vertical or horizontal movements were extremely rare. Three such periods of vertical movements were seen. After each of these the dream content involved a predominance of action in the vertical plane One *S* dreamed of standing at the bottom of a tall cliff operating some sort of hoist and looking up at climbers at various levels and down at the hoist machinery. Another *S* dreamed of climbing up a series of ladders looking up and down as he climbed. In the third instance the dreamer was throwing basketballs at a net, first shooting and looking up at the net, and then looking down to pick another ball off the floor. Only one instance of pure horizontal movement was seen. In the associated dream *S* was watching two people throwing tomatoes at each other. On 10 occasions *S*s were awakened after 1 min. of little or no eye movement In these, the dreams all had the common property that the dreamer was watching something at a distance or just staring fixedly at some object In two of these awakenings in different *S*s the patterns were the same, as follows: about a minute of ocular inactivity followed by several large movements to the left just a second or two before the awakening. Both instances, interestingly enough, were virtually identical as regards dream content. In one case *S* was driving a car and staring at the road ahead. He approached an intersection and was startled by the sudden appearance of a car speeding at him from the left as the bell rang. In the other, the dreamer was also driving a car and staring at the road ahead. Just before the awakening he saw a man standing on the left side of the road and hailed him as he drove past.

In the 21 awakenings after a mixture of movements *S*s were always looking at things close to them, objects or people. Typical reports were of talking to a group of people, looking for something, fighting with someone, and so forth. There was no recall of distant or vertical activity.

In order to confirm the meaningfulness of these relationships, 20 naive *S*s as well as 5 of the experimental *S*s were asked to observe distant and close-up activity while awake. Horizontal and vertical electrodes were attached. The eye-movement potentials in all cases were comparable in both amplitude and pattern to those occurring during dreaming. Furthermore, there was virtually no movement, as indicated by the eye potentials, when viewing distant activity, and much movement while viewing close-up activity. Vertical eye-movement potentials were

always at a minimum except for the upward movements accompanying blinking, and in a few cases when *E* tossed a ball in the air for them to watch.

Discussion

The results of these experiments indicate that dreaming accompanied by REM's and a low-voltage electroencephalogram occurred periodically in discrete episodes during the course of a night's sleep. It cannot be stated with complete certainty that some sort of dream activity did not occur at other times. However, the lack of recall and also the fact that the brain waves were at the lightest level of sleep only during REM periods and at deeper levels at all other times, makes this unlikely. The few instances of dream recall during NREM periods are best accounted for by assuming that the memory of the preceding dream persisted for an unusually long time. This is borne out by the fact that most of these instances occurred very close, within 8 min., after the end of REM periods.

Other workers have attempted to relate dreaming to physiological phenomena during sleep. Wada (**12**) felt that dreaming and gastric contractions occurred simultaneously. However, this conclusion was based on only seven awakenings in two *S*s. One was unable to recall dream content although he felt he had been dreaming and the other remembered dream content in 3 of 4 awakenings. Scantlebury, Frick, and Patterson (**11**) also studied gastric activity and dreaming. They felt, on the basis of three instances of dream recall out of seven awakenings, that the two were probably related, but judiciously stated that "the exact time during which a dream occurs is elusive of record." The occurrence of dreaming during a series of foot twitches occurring immediately after the onset of sleep was postulated by McGlade (**9**). However, he based this conclusion mainly on dreams recalled on the morning after the experiments which is highly unreliable, and only 3 out of the 25 *S*s studied exhibited foot twitches.

Incidental observations have been made on the occurrence of dreaming by investigators studying brain waves during sleep (**2, 3, 6, 7, 8**). All stages of brain waves were related to dreaming in these five papers, but no mention was made of whether or not actual dream content was recalled, and the number of reports by sleepers was generally very small.

In other studies of dreaming, excellently reviewed by Ramsey (**10**), attempts were made to localize dream activity by simply awakening *S*s at various times during the night. In general it was found that dreams might be recalled at any time during the night, but that most were recalled in the later hours of sleep. This would correspond to the statistical incidence of REM's as previously reported (**1, 4**), and is also consistent with the finding in this study that, even when the awakenings occurred during REM periods, recall was still more difficult earlier in the night.

It was stated herein that all *S*s showed periods of REM's *every* night they slept. This was also the case in another briefly reported series of experiments involving 16 *S*s who were observed a total of 43 nights (**5**). It is felt on the basis of these and other studies which are unreported that periods of REM's and dreaming and the regularity with which they occur are an intrinsic part of normal sleep. In view of this, the failure to observe REM's in occasional *S*s reported in earlier work (**1, 4**) deserves some consideration. One explanation is that the recording was done by sampling rather than continuously. If the REM periods were shorter than usual, they may have occurred in the intervals between the samples, thus escaping observation. Another explanation is that a lower amplification of the REM potentials was employed which, although usually adequate, did not clearly record very small movements. A third possibility is that the dreams of these *S*s happened to be the sort, such as watching

distant activity, in which eye movement was at a minimum. Since the association of the characteristic low-voltage, non-spindling EEG was not realized at the time and thus could not aid in identifying this sort of period, they very likely would have been overlooked.

There was nothing in the experiments reported in this paper to indicate that the dreams occurred instantaneously, or with great rapidity, as some have supposed. Rather, they seemed to progress at a rate comparable to a real experience of the same sort. An increment in the length of REM periods was almost invariably associated with a proportional increase in the length of the dream. This could not have occurred if dreaming were instantaneous, since any length of REM periods would then easily accommodate a virtually infinite amount of dream activity.

It seems reasonable to conclude that an objective measurement of dreaming may be accomplished by recording REM's during sleep. This stands in marked contrast to the forgetting, distortion, and other factors that are involved in the reliance on the subjective recall of dreams. It thus becomes possible to objectively study the effect on dreaming of environmental changes, psychological stress, drug administration, and a variety of other factors and influences.

Summary

Regularly occurring periods of REM's were observed during every night of experimental sleep in nine adult *Ss*. A high incidence of dream recall was obtained from *Ss* when awakened during REM periods and a very low incidence when they were awakened at other times. A series of awakenings was done either 5 or 15 min. after the REM's (dreaming) had begun and *Ss* judged the correct dream duration with high accuracy. The pattern of the REM's was related to the visual imagery of the dream, and the eye movements recorded in analogous situations while awake corresponded closely in amplitude and pattern to those observed during dreaming.

References

1. Aserinsky, E., & Kleitman, N. Two types of ocular motility occurring in sleep. *J. appl. Physiol.*, 1955, **8,** 1–10.
2. Blake, H., Gerard, R., & Kleitman, N. Factors influencing brain potentials during sleep. *J. Neurophysiol.*, 1939, **2,** 48–60.
3. Davis, H., Davis, P., Loomis, A. L., Harvey, E. N., & Hobart, G. A. Human brain potentials during the onset of sleep. *J. Neurophysiol.*, 1938, **1,** 24–38.
4. Dement, W. Dream recall and eye movements during sleep in schizophrenics and normals. *J. nerv. ment. Dis.*, 1955, **122,** 263–269.
5. Dement, W., & Kleitman, N. Incidence of eye motility during sleep in relation to varying EEG pattern. *Fed. Proc.*, 1955, **14,** 216.
6. Henry, C. E. Electroencephalographic individual differences and their constancy. I. During sleep. *J. exp. Psychol.*, 1941, **29,** 117–132.
7. Knott, J. R., Henry, C. E., & Hadley, J. M. Brain potentials during sleep; a comparative study of the dominant and non-dominant alpha groups. *J. exp. Psychol.*, 1939, **24,** 157–168.
8. Loomis, A. L., Harvey, E. N., & Hobart, G. A. Cerebral states during sleep as studied by human brain potentials. *J. exp. Psychol.*, 1937, **21,** 127–144.
9. McGlade, H. B. The relationship between gastric motility, muscular twitching during sleep and dreaming. *Amer. J. digest. Dis.*, 1942, **9,** 137–140.
10. Ramsey, G. Studies of dreaming. *Psychol. Bull.*, 1953, **50,** 432–455.
11. Scantlebury, R. E., Frick, H. L., & Patterson, T. L. The effect of normal and hypnotically induced dreams on the gastric hunger movements of man. *J. appl. Psychol.*, 1942, **26,** 682–691.
12. Wada, T. An experimental study of hunger and its relation to activity. *Arch. Psychol.*, N. Y., 1922, **8,** No. 57.

(Received May 9, 1956)

Journal of Experimental Psychology
Vol. 53, No. 5, 1957

EFFECT OF IRRELEVANT THIRST MOTIVATION ON A RESPONSE LEARNED WITH FOOD REWARD

G. ROBERT GRICE AND JOHN D. DAVIS

University of Illinois

The generalized response-activating character of drive (*D*) which has been hypothesized by Hull (**6**) and Spence (**13**) has received a fairly substantial amount of empirical support. Amsel (**1**) found that rats will perform a running habit motivated by fear more vigorously when they are also hungry. Amsel and Maltzman (**2**) found that the presence of emotionality will, under certain circumstances, increase the amount of drinking behavior. The Iowa studies of anxiety in human *Ss* have been based on this assumption and, in general, have yielded results in accord with prediction (e.g., **5, 12, 16**). In the case of purely appetitional drives, Webb (**18**) trained rats on a panel-pushing response under food deprivation for food reward and conducted extinction trials under various degrees of water deprivation with the hunger drive satiated. He found that the number of extinction responses was an increasing function of the number of hours of water deprivation. This result was confirmed by Brandauer (**3**) in a similarly designed experiment employing a bar-pressing response.

Verplanck and Hayes (**17**) have raised a serious question concerning the interpretation of the Webb experiment in terms of nonspecific activation by a generalized drive state. Their question concerns the possibility of the independent manipulation of the hunger and thirst drives. They found that rats deprived of water with food continuously available and then allowed to drink would now eat an amount almost equivalent to the eating produced by an equal period of food deprivation. They found, further, that water-deprived rats eat only about 60% as much food as rats not so deprived. Their findings were also similar for the reverse drive arrangement. The implication of these findings for studies like Webb's is that the hunger drive may actually be present in considerable strength following water deprivation. Possibly adequate hunger satiation is impossible because of the reduced eating by thirsty animals. Webb was also aware of this possible interpretation of his findings, but tentatively rejected it in favor of "Hull's hypothesis of 'a generalized drive state' being 'contributed to' by the irrelevant drives (defined by deprivation time) operative at the time" (**18,** p. 13).

The present experiment is an attempt to investigate these alternative explanations of the role of an irrelevant thirst drive. Four groups of rats were trained on a panel-pushing response for food reward under hunger motivation. One control group was then given extinction trials under hunger motivation and a second while satiated with both food and water. One experimental group was tested under thirst conditions. This much constitutes the basic design of the Webb study. A second experimental group was subjected to the same water deprivation and food satiation procedures as the thirsty group, but was then allowed to drink to satiation prior to the test. The

most obvious implication of the generalized drive hypothesis is that, with the irrelevant drive satiated, this group should be lower in response strength than the thirsty group. The implication of the Verplanck and Hayes finding, however, is that this group should respond more than the thirsty group, or at least be equal to it.

Method

Subjects

The *S*s were 80 experimentally naive, albino rats of Sprague-Dawley strain. There were 10 males and 10 females in each of the four experimental groups. Their ages ranged from 70 to 90 days at the beginning of the experiment.

Apparatus

The apparatus consisted of a gray plywood box 12 in. long, 10 in. wide, and 6 in. high, inside dimensions. The top was of hinged lucite and there was a row of thirty-five $\frac{3}{16}$-in. holes $4\frac{1}{2}$ in. above the floor extending around the walls to permit the passage of air. The apparatus was illuminated by a 15-w. bulb and reflector 24 in. above the floor at the rear of the box. At the middle of one end there was a lucite panel $1\frac{1}{2}$ in. wide by 6 in. high and hinged at the top. When closed the panel was flush with the end of the box. A vertical sliding door of $\frac{1}{4}$-in. masonite and painted the same as the box was placed immediately in front of the panel. With the door raised the panel was exposed to a height of $2\frac{1}{4}$ in. When the door, which was controlled manually by means of a string and counterweight arrangement, reached a height of 1 in., a holding relay was closed and a Standard Electric Timer was started. When the panel was pushed outward a distance of $\frac{1}{2}$ in. at its bottom, the circuit was broken by means of a tilting mercury switch, and the clock was stopped. Food pellets, consisting of Tail Curler Junior and weighing approximately .05 gm., were placed in a slight depression in the floor immediately behind the panel. It was impossible for the rat to obtain the pellet without stopping the clock.

Preliminary Training

Throughout training *S*s were maintained on a diet of 9 gm. of Purina Laboratory Chow per day, administered in individual feeding cages and subject to a 1-hr. maximum eating period. Water was continuously available. They were adapted to this schedule for a minimum of seven and a maximum of nine days prior to preliminary training. On each of the last two days of the adaptation period, *S*s received 10 pellets of the type used in training along with the daily ration. During preliminary training and training, *S*s were fed immediately after the experimental session. Since, after adaptation, *S*s actually ate in about 30 min., the hunger drive was produced by approximately $23\frac{1}{2}$ hr. of deprivation. On the first pretraining day, 10 pellets were placed in the panel opening and *S*s were allowed to eat them with the sliding door and the panel remaining open. On the second pretraining day, a single pellet was placed in the food hole with the panel open and the door closed. The door was raised and remained open until *S* had obtained and eaten the pellet. This was repeated until *S* had consumed 10 pellets.

Panel Training

On the next day, a single pellet was placed in the apparatus with both the panel and door closed. The door was raised and remained open until *S* had obtained the pellet by pushing the panel open with its nose. The door was then closed carefully to avoid frightening the animal. Latency was recorded to 1/100 sec. On the first training day only, *E* held the panel open with a string after the clock was stopped until *S* obtained the pellet on those few trials when it was necessary. Fifteen trials were presented with an intertrial interval of 15 sec. timed from the lowering of the door at the end of one trial until it was raised at the beginning of the next. On each of the next two days, 15 additional trials were presented, making a total of 45 acquisition trials.

Test Day

On the following day, 30 trials were presented under conditions like that during acquisition except that no food was presented. The criterion of the occurrence of a response was the pushing of the panel far enough to stop the clock. Whenever *S* failed to respond in 30 sec., the door was lowered and the next trial was presented after the intertrial interval. Such a trial was counted as no response and as a maximum latency.

The procedures used to produce the different drive conditions in the four experimental groups were as follows:

Hungry Group.—This group was tested under the same food deprivation conditions which prevailed during training.

Satiated Group.—For this group, water and a supply of Purina Laboratory Chow were con-

tinuously available during the 24-hr. period preceding the test. One and one-half hours before the test trials, additional Purina Chow and a rat which had been deprived of food for 48 hr. were introduced into *S*'s living cage; 1 hr. before the test, *S* was placed in an individual feeding cage with water, Purina Chow, and a dish of the pellets used as incentive in the apparatus; and 30 min. before the test *S* was placed in an individual feeding cage with water available, but no food. This 30-min. period was introduced in order to make the group comparable to the Thirsty-Drink Group with respect to the availability of food.

Thirsty Group.—This group was treated in the same way as the Satiated Group, except that water was removed 22 hr. before the test, and no water was available during the special food satiation procedures. For 30 min. prior to the test, *S* was placed in an individual feeding cage with no food or water. Water deprivation of 22 hr. was selected because it was Webb's most effective irrelevant drive condition.

Thirsty-Drink Group.—The treatment of this group was identical in every way to that of the Thirsty Group, except that a water bottle was available in the individual feeding cage for the 30 min. prior to the test. It was noted that all *S*s drank vigorously, but had ceased by the end of 30 min.

Results

The results of the test day are presented in terms of three response measures—the number of responses during the 30 test trials, a speed measure, and a rate of responding measure. The speed measure was obtained by finding the median latency in seconds of the first 10 test trials for each *S*. This median was transformed to its reciprocal and multiplied by 1000. For the rate measure, the total amount of time during which the panel was available to each *S* was calculated by adding the latencies, including the 30 sec. failures, of the 30 trials. The number of responses was then divided by the total time in minutes, and the result is expressed as the number of responses per minute. This measure is novel in this type of situation but has the advantage of utilizing all of the data collected during the test trials. The medians, means and *SD*'s of the three measures for the four groups are presented in Table 1. Since several of the distributions were markedly skewed for all measures, the medians are regarded as the better index of central tendency.

In order to evaluate the over-all effects of the drive conditions, sex, and possible interactions, Wilson's nonparametric test (**20**) which permits such an analysis was applied to the number-of-response data. The results of this test are summarized in Table 2. As may be seen, the effect produced by the drive variable was highly significant, while the Sex and Sex × Drive interaction effects were not significant.

The chief interest in the study, of course, is not in over-all significance, but in specific comparisons between conditions. These comparisons were made by means of the Mann and

TABLE 1

Results of Test Day

(N = 20 in each group)

Group	No. of Responses			Speed			Responses/Min.		
	Median	Mean	*SD*	Median	Mean	*SD*	Median	Mean	*SD*
Hungry	26.50	24.90	4.18	246	319	232	4.68	5.75	4.30
Satiated	13.50	15.20	7.33	105	150	140	1.29	2.51	3.64
Thirsty	13.00	13.70	7.38	71	101	87	1.18	1.60	1.36
Thirsty-Drink	22.00	20.75	6.37	218	227	157	3.62	3.87	2.91

TABLE 2

NONPARAMETRIC ANALYSIS OF NUMBER OF RESPONSES

Source	*df*	χ^2
Drive	3	18.85*
Sex	1	.46
D × S	3	5.81
Total	7	25.12*

* $p < .001$.

Whitney *U* test. For the number-of-responses measure, the Hungry Group differed from the Satiated and Thirsty Groups at the .001 level and from the Thirsty-Drink Group at the .03 level. The Thirsty-Drink Group differed significantly from the Satiated Group at the .02 level and from the Thirsty Group at the .01 level. The Satiated and Thirsty Groups did not differ significantly. The results of this test when applied to the other response measures were essentially the same, except that the difference between the Hungry and Thirsty-Drink Groups was not significant for the other measures.

An examination of the group measures in Table 1 and the significance values will immediately indicate that the results of the test day were quite clear. As was to be expected, the Hungry Group showed the greatest response strength. There was no significant difference between the Satiated and Thirsty Groups, although for all measures the Satiated Group showed a small superiority. Thus, the present study fails to confirm the findings of Webb and Brandauer who found the irrelevant thirst drive groups superior to the satiated control. The satiation of the thirst drive in the Thirsty-Drink Group not only did not reduce the level of response, but actually increased it as indicated by this group's significant superiority over the Thirsty Group.

DISCUSSION

In considering the failure of the present experiment to confirm Webb's finding, it should be clearly understood that this experiment was not, strictly speaking, a repetition of the Webb study since the apparatus and experimental procedures were not the same. However, the rationale of the design was the same and there is no apparent reason why it should not provide an equally good test of the generalized drive hypothesis. The hypothesis should lead to the same prediction in both cases. This negative finding may indicate a serious restriction upon its generality, and we must at least state that a response learned with food reward will not always be activated by an irrelevant thirst drive. In view of Brandauer's confirmation of Webb's result, as well as the nature of inference concerning the null hypothesis, the weight of the evidence remains on the positive side, but further clarification of the problem must await future research.

More can be said in interpretation of the finding that allowing the water-deprived *S*s to drink prior to the test resulted in an increase in response strength. The fact that reducing the thirst drive produced more vigorous responding implies that the presence of the drive was actually depressing the response. This is, of course, the direct opposite of the drive summation theory and of the interpretation of Webb. There are various ways in which this depressing effect might come about. One possible explanation is in terms of the concept of incentive motivation (*K*) as developed by Spence (**14**). It may be that the anticipatory goal response for food ($r_{g\text{food}}$), the mechanism which Spence assumes to provide the basis for *K*, is eliminated or seriously reduced by the presence of the thirst drive, but returns following satiation for water. This explanation is not entirely satisfactory, however, since there is strong indication in several maze-learning ex-

periments (**7, 11, 15**) that r_{gfood} does occur in the presence of the thirst drive—particularly if it is not closely associated with drinking responses.

Another possible explanation of the depression effect is suggested by the work of Kendler and his associates (**8, 9, 10**). They have presented evidence that rats which have learned to go to food on one side of a T-maze show a tendency to avoid the food side when made thirsty. This is explained by hypothesizing that fractional anticipatory eating responses and the condition of water deprivation are incompatible, producing a conflict when they are simultaneously active. This state of conflict is assumed to be reduced by the avoidance of cues evoking the anticipatory eating response. This hypothesis seems to account for the present finding rather well. There does, however, appear to be some difficulty in completely reconciling this theory with the fact that both cues associated with food (**4**) and the sight of food (**19**) will produce secondary reinforcement in thirsty rats.

The question now to be considered is that of the drive state responsible for the increased level of response in the Thirsty-Drink Group. The prediction made on the basis of the Verplanck and Hayes analysis was clearly supported. It appears that the hunger drive may have been present in considerable degree even though responding was depressed by the thirst drive in the Thirsty Group. Assuming that the Webb and the Brandauer results are reproducible at least under some conditions, it now seems likely that the positive relation between response strength and hours of water deprivation was not the result of activation by the irrelevant thirst drive. The present results strongly suggest the implication that the hunger drive was also increasing during water deprivation and that the motivation may have been entirely "relevant."

One possibility, not yet considered, is that water deprivation actually reduces the hunger drive, although presumably not the need state itself. However, considering the increasing function found by both Webb and Brandauer together with the present evidence that the effect of thirst is depressing, the best guess at present seems to be that hunger does increase during water deprivation but that performance is inhibited by some mechanism such as the ones suggested here.

At the present time, it is probably wise to be cautious in evaluating the generalized drive theory in the light of these data. While this study does not support the theory and further implies that the Webb and Brandauer studies do not necessarily support it either, there is enough additional evidence to suggest that it has considerable validity. The theory may be valid as a general principle, but subject to restrictions arising from interaction effects between some specific drive states.

Summary

Four groups of rats were trained under hunger motivation to push a panel for food reward. One group was given extinction trials under hunger motivation, one while satiated, and one under thirst motivation. The fourth group was deprived of water like the thirst group, but was allowed to drink before the extinction trials.

There was no significant difference in the extinction measures between the satiated and thirsty groups. Thus, the experiment failed to confirm previous studies showing an activating effect of irrelevant thirst motivation.

The water-deprived group which was allowed to drink displayed more response strength during extinction than the thirsty group. Possible reasons for a depressing effect of the thirst drive are discussed.

The results are interpreted as implying that the positive relation previously found between hours of water deprivation and the strength, during extinction, of a response learned with food reward may be due to increasing strength of the hunger drive.

References

1. Amsel, A. The combination of a primary appetitional need with primary and secondary emotionally derived needs. *J. exp. Psychol.*, 1950, **40**, 1–14.
2. Amsel, A., & Maltzman, I. The effect upon generalized drive strength of emotionality as inferred from the level of

consummatory response. *J. exp. Psychol.*, 1950, **40**, 563–569.

3. Brandauer, C. M. A confirmation of Webb's data concerning the action of irrelevant drives. *J. exp. Psychol.*, 1953, **45**, 150–152.
4. D'Amato, M. R. Transfer of secondary reinforcement across the hunger and thirst drives. *J. exp. Psychol.*, 1955, **49**, 352–356.
5. Farber, I. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.
6. Hull, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
7. Kendler, H. H. The influence of simultaneous hunger and thirst drives upon the learning of two opposed spatial responses of the white rat. *J. exp. Psychol.*, 1946, **36**, 212–220.
8. Kendler, H. H., Karasik, A. D., & Schrier, A. M. Studies of the effect of change of drive: III. Amounts of switching produced by shifting drive from thirst to hunger and from hunger to thirst. *J. exp. Psychol.*, 1954, **47**, 179–182.
9. Kendler, H. H., & Levine, S. Studies of the effect of change of drive: I. From hunger to thirst in a T-maze. *J. exp. Psychol.*, 1951, **41**, 429–436.
10. Kendler, H. H., Levine, S., Altchek, E., & Peters, H. Studies of the effect of change of drive: II. From hunger to different intensities of a thirst drive in a T-maze. *J. exp. Psychol.*, 1952, **44**, 1–3.
11. McAllister, W. R. The spatial relation of irrelevant and relevant goal objects as a factor in simple selective learning. *J. comp. physiol. Psychol.*, 1952, **45**, 531–537.
12. Montague, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.
13. Spence, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. *Learning theory, personality theory, and clinical research. The Kentucky symposium*. New York: Wiley, 1954.
14. Spence, K. W. Theoretical interpretations of learning. In C. P. Stone (Ed.), *Comparative psychology*. (3rd ed.) New York: Prentice-Hall, 1951.
15. Strange, J. R. Latent learning under conditions of high motivation. *J. comp. physiol. Psychol.*, 1950, **43**, 194–197.
16. Taylor, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
17. Verplanck, W. S., & Hayes, J. R. Eating and drinking as a function of maintenance schedule. *J. comp. physiol. Psychol.*, 1953, **46**, 327–333.
18. Webb, W. B. The motivational aspect of an irrelevant drive in the behavior of the white rat. *J. exp. Psychol.*, 1949, **39**, 1–14.
19. Wike, E. L., & Casey, A. The secondary reinforcing value of food for thirsty animals. *J. comp. physiol. Psychol.*, 1954, **47**, 240–243.
20. Wilson, K. V. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, **53**, 96–101.

(Received May 10, 1956)

Journal of Experimental Psychology
Vol. 53, No. 5, 1957

PREDICTION FROM AND INTERACTION AMONG MULTIPLE CONCURRENT DISCRIMINATIVE RESPONSES [1]

CHARLES W. ERIKSEN [2]

The Johns Hopkins University

In the present study *O*s were presented with a size-judgment task under the method of absolute judgment. Instead of using only one response or judgment to each stimulus presentation as is usually done, the present experiment varied the number and the nature of the judgmental responses. Here *O*s were required to make not only a verbal judgment to each stimulus but were also required to move a lever to a distance proportionate to the size of the presented stimulus and to make a "second guess" verbal judgment.

The primary aim of this study was to examine the role of supplemental responses in a situation that contained a fair amount of inherent uncertainty. Specifically, this experiment asked whether there is an independent relation between each of the several response systems and the stimulus or whether all of the responses will function essentially as a unit. For example, the first verbal judgment may serve to classify the perception for *O*, and the lever response and the second verbal judgment may then be mere reflections of the first judgment without any independent relation to the stimulus. Phrased statistically, the question becomes one of determining whether there exist significant partial correlations between each of the response systems and the stimulus when the remaining response systems are held constant.

A second question that can be asked concerns the possibility of differential accuracy between the verbal judgment and the lever-movement response. Verbal responses in general, and in this experiment in particular, are of a discrete categorizing nature. On the other hand, the linear movement of a lever is, for practical purposes, continuous. In the present experiment an *O* who is confronted with Stimulus X may be uncertain as to whether this is Stimulus X or X + 1. For his verbal judgment he must respond with either the response for X or the response for X + 1, but his lever movement is not so restricted. If the standard lever movement in response to X was in the range of 8–10 mm., one might reasonably expect the movement to be closer to 10 mm. when *O* was uncertain as to whether the stimulus was X or X + 1 and closer to 8 mm. when *O*'s uncertainty was between X and X − 1. The continuous character of the lever-movement response may thus contribute to a higher stimulus-response correlation for the lever response than would be obtained for the verbal responses, as well as contributing to a partial correlation between the stimuli and the lever response by reflecting the direction of *O*'s uncertainty.

It should be noted that the present experiment also has an important relation to the subception effect (6). It has been shown elsewhere (1) that

[1] This investigation was supported by Research Grant M-1037 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service.

[2] Now at the University of Illinois.

the subception experiment of Lazarus and McCleary (6) is basically a situation requiring multiple responses (a GSR and a verbal response) to the same stimulus, and the subception effect consists of a significant partial correlation between the stimulus and the GSR when the verbal responses are held constant. That such a partial correlation does exist has been further shown by Eriksen (2), and it was partly the purpose of the present study to increase the understanding of the circumstances under which such partial correlations will be obtained.

Method

Design.—The experimental task consisted of the absolute judgment of a series of 11 different sized, square stimuli. Three experimental groups were employed. In Group A, *Os* were required to use verbal responses consisting of the numbers "1" through "11" where the number "1" was assigned as the response to the smallest size, number "11" to the largest size, and the intermediate numbers were assigned to the intermediate stimuli in order of increasing size. In addition to the verbal judgment, *Os* were asked to move a vertically mounted lever through a segment of arc that was proportionate to the size of the stimulus. After the verbal judgment and the lever response to each stimulus, each *O* was also asked to make a second verbal judgment of the stimulus, this time assuming that he had been told that his first verbal judgment had been wrong.

The experimental treatment was the same for *Os* in Group B except for the manner in which they made the lever response. In Group A, *Os* had the lever in full view so that they were able to use visual cues while moving the lever. In Group B, *Os* had to reach their hands through a cloth curtain and position the lever without the use of visual control and feedback. This condition was employed as a partial control for Group A since it was anticipated that the lack of visual monitoring would result in greater error in the lever response which in turn should reduce the possibility of partial correlations between the stimuli and the lever responses.

The *Os* in Group C differed from the other two groups in that they were permitted to use only two verbal responses, "Yes" and "No." During the pretraining series that all groups were given, the middle-sized stimulus was presented 20 times. During the experimental sessions, *Os* in Group C were asked to report "Yes" if the stimulus presented was the same size as the training stimulus and to report "No" if it was larger or smaller. The movement response for this group was the same as for *Os* in Group A, but no "second guess" verbal response was permitted them.

Materials.—The *O* was seated across a table from *E*. A wooden partition separated and obscured *E* from *O*. Slightly to the left of *O* and approximately at eye level was an 8 × 8 in. aperture containing a milk-glass screen on which the size stimuli were presented. Projecting from the table top directly in front of *O*'s right shoulder was an 8-in. knobbed lever which could be moved through 60 degrees of arc from *O*'s left to right. This gave a total length of 8.38 in. or 21.28 cm. Through a system of strings and pulleys, the movement of the lever was recorded on a kymograph on *E*'s side of the partition. The arrangement of the table and the positioning of the lever were such that *O* was unable to support his elbow while positioning the lever. For Group B, the lever was covered with a cardboard shield so that *Os* had to reach their hands through a cloth curtain in order to position the lever.

The size stimuli were constructed from squares of dark gray paper photographed onto 35-m. slides. They were projected onto the milk-glass screen. The projected size of the smallest square was 4 mm. square, and the stimuli increased in size by steps of 2 mm. per side so that the largest square was 34 mm. square.

Procedure.—All *Os* were run individually. Prior to beginning the experimental judgments, *Os* were given a preliminary practice session which consisted of 20 presentations of the mid-sized stimulus during which *Os* practiced positioning the lever to the midposition. Following this practice, each of the 11 size stimuli was presented four times in a random sequence with the restriction that each size would occur twice in the first 22 presentations. The lever positioning response always began with the lever in the extreme left position, and *O* was required to move it toward his right a distance proportionate to the size of the stimulus. He was further instructed to make only a slight movement for the smallest stimulus and to make the movement for the largest stimulus very near to the maximum.

Subjects.—A total of 31 *Os* were run, 10 in Group A, 12 in Group B, and 9 in Group C. All were male undergraduate students drawn from the elementary psychology course at The Johns Hopkins University. They were randomly assigned to the three experimental groups.

TABLE 1

AVERAGE INTERCORRELATION BETWEEN STIMULI AND RESPONSE SYSTEMS AND AMONG RESPONSE SYSTEMS FOR GROUPS A AND B

Variables	Groups		t*
	A	B	
r_{SV_1}	.957	.954	.45
r_{SV_2}	.935	.933	.22
r_{SL}	.952	.917	3.14
$r_{V_1V_2}$	.939	.957	2.03
r_{V_1L}	.985	.959	4.20
r_{V_2L}	.933	.920	1.60

* With 19 *df*, *t* values of 2.09 and 2.86 are significant at the .05 and the .01 level, respectively.

RESULTS

The first step in the analysis of the data was to determine the average intercorrelation between the stimuli and the different response systems for each of the experimental groups. For each *O* the product-moment correlation between the stimuli (S) and the first verbal response (V_1), the second verbal response (V_2), and the lever response (L) was computed, as well as the intercorrelations among these three response systems. These correlation coefficients obtained from the individual *O*s were then transformed to Fisher's z and the average z value computed by experimental groups for each of the six stimulus-response and response-response correlations. The significance of differences between Groups A and B were evaluated by computing the variance of the obtained z distributions and using a *t* test for the differences between the mean z values. The resulting values of *t* are shown in Table 1 along with the average intercorrelations for Groups A and B.[3] The average intercorrelations in this table were obtained by reconverting the mean z values into correlation coefficients.

As can be seen from the table, the average intercorrelation between the stimuli and the first and second verbal responses (r_{SV_1} and r_{SV_2}) does not differ significantly between Groups A and B. However, the correlation between the stimuli and the lever responses (r_{SL}) is significantly larger in Group A, as is also the correlation between the lever and the first verbal responses (r_{V_1L}). In Group B, on the other hand, the correlation between the two verbal responses ($r_{V_1V_2}$) tends to be higher although this difference is not quite significant at the .05 level.

These group differences in average correlation are for the most part to be expected due to the greater error in the lever responses of Group B. Since *O*s in this group were required to position the lever without the advantage of visual monitoring of the response, the precision of their settings was less than for Group A. This greater error in the lever response would be expected to reduce the correlation of this response with the stimuli as well as with the other responses. This interpretation is further supported when we examine the average correlation between the stimuli and the lever response for Group C. Like Group A, the *O*s in Group C were allowed to visually monitor their lever responses, although their verbal responses were limited to "Yes" and "No." The average correlation between the stimuli and the lever responses for this group was .949 which is significantly larger than the corresponding value of .917 obtained for Group B ($t = 2.97$ with 20 *df*) although not

[3] Group C is not represented in this table since the only product-moment correlation that could be computed for *O*s in this group was that between the stimuli and the lever response. The average value of this correlation was .949.

significantly nor appreciably different from the value for Group A.

When differences within experimental groups are examined, it is found that in both Groups A and B the first verbal response system shows the highest average correlation with the stimuli. In Group A the lever responses have a higher correlation with the stimuli than do the second verbal responses, but this difference is reversed in Group B. These differences in magnitude of average correlation between the stimuli and the three response systems are all significant beyond the .05 level for Group B when evaluated by the *sign test*, but in Group A only the difference between the intercorrelation of the first and second verbal responses with the stimuli is significant at this level.[4]

While the above analysis shows that all three of the response systems are highly correlated with the stimuli, the major interest of this study is the use that *O*s make of the additional responses or judgments that are permitted them. For example, if *O* makes an error on his first verbal judgment, do both of the other responses reflect this error or do they tend to correct for it? Alternatively stated, do each of the response systems provide information about which of the stimuli has occurred over and above the information provided by either or both of the other response systems? These questions can be answered by examining the correlations between the stimuli and a response system with either or both of the remaining response systems partialed out.

The partial correlation between the stimuli and each of the three response systems was computed for each *O* in Groups A and B. The obtained partial correlations were then transformed to Fisher's z and averaged by experimental groups. To determine whether a statistically reliable partial correlation existed within an experimental group, the variance of the obtained z distribution for this partial correlation was computed and a *t* test applied to the average z value to determine whether it was significantly greater than zero. Between-group differences were also evaluated in this manner, but here the *t* test was applied to the difference between mean z values for the two groups.

Table 2 summarizes the results of these analyses. The average partial correlations shown in the table were obtained by transforming the corresponding mean z values back to

TABLE 2

AVERAGE PARTIAL CORRELATIONS BETWEEN STIMULI AND RESPONSE SYSTEMS FOR GROUPS A AND B

Partial r	Groups		t†
	A	B	
$r_{SV_1 \cdot V_2}$	.647**	.578**	.11
$r_{SV_1 \cdot L}$	.397**	.670**	5.73
$r_{SV_2 \cdot V_1}$	.515**	.282*	1.83
$r_{SV_2 \cdot L}$	.430**	.591**	2.82
$r_{SL \cdot V_1}$	.273*	.011	2.20
$r_{SL \cdot V_2}$	.635**	.405**	3.01
$r_{SV_1 \cdot V_2 L}$	.300**	.454**	2.16
$r_{SV_2 \cdot V_1 L}$	.397**	.345**	.55
$r_{SL \cdot V_1 V_2}$	.309*	.060	1.81

† With 19 *df*, *t* values of 2.09 and 2.86 are significant at the .05 and the .01 level, respectively.
* Significantly different from zero at the .05 level.
** Significantly different from zero at the .01 level.

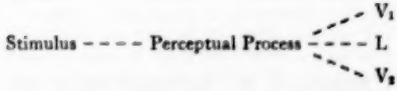
[4] Since the z values representing the intercorrelations for the different *O*s within experimental groups are intercorrelated in turn in a highly complex way, the significance of the differences between the correlations of the different response systems with the stimuli could not be evaluated by the usual *t* test for correlated means. Instead, the sign test was applied by determining the number of *O*s in each group for whom the correlation being tested was higher than the correlation with which it was being compared. Thus, for example, in Group B, r_{SV_1} was greater than r_{SL} for each of the 12 *O*s in the group ($P < .01$).

correlation coefficients, and the asterisks are used to designate those average partial correlations that are significantly greater than zero. The significance of the between-group differences are indicated by the accompanying values of *t*.

For Group A all nine of the partial correlations have an average value significantly greater than zero, indicating that each of the three response systems has a relationship with the stimuli independent of either or both of the other two response systems. For Group B, however, only the first and the second verbal responses have an independent relationship with the stimuli. The failure of the lever responses to show an independent relationship with the stimuli in this group can be attributed to the greater error in this response system arising from the lack of visual monitoring of these responses. This greater error in the lever responses would also seem to account for the differences in average partial correlation between the two groups.

Discussion

The above results can be efficiently summarized and discussed in terms of a schematic model of perception that has been proposed elsewhere by Eriksen (1) and Garner, Hake, and Eriksen (5). Briefly, the model provides for a distinction between perception as a hypothetical concept and the responses by which it is inferred. In keeping with this distinction we can schematize the present experiment along the following lines:



Here the three response systems available to Groups A and B are conceived as essentially independent of each other, each responding to the perceptual process and each having the opportunity to be influenced by unique sources of error.

Such a schema provides the necessary and sufficient conditions for the present experimental results. Considering the different response systems as independent provides for the result that each response system has an independent relation to the stimulus or perceptual process. By further conceiving that each response system can be uniquely affected by certain variables or sources of error, provision is made for the significant partial correlations between the stimuli and the different responses.

In addition to offering an efficient conceptualization of the present results, this model has the advantage of focusing attention upon the variables that will lead to noncorrelated errors in the different response systems. While neither the model nor the experimental data require that all the error in a response system be unique to that system, the significant partial correlations that were obtained do require that each of the three response systems have a source of error that is not common to either of the other two. Some factors leading to these noncorrelated errors have already been indicated. For example, it has been shown that the number of alternative responses in a verbal response system will affect the kind and number of errors in this system without affecting a concurrent autonomic response (2), and there is reason to believe that unequal habit strengths among verbal responses might have a similar effect (2, 4).

In the present study it was found that increasing the error in the lever response through elimination of visual monitoring did not affect the precision of the concurrent verbal judgments. While the correlation between the stimuli and the lever response was significantly less in Group B, there was no significant difference in the accuracy with which Groups A and B used the verbal responses. Similarly, there was no indication that the precision of the verbal responses had any effect upon the accuracy of the lever response. The *Os*

in Group C with only two verbal labels to classify the stimuli were as accurate in their lever responses as the Group A *Os* who were allowed to use 11 different verbal labels to classify the stimuli. However, the possibility cannot be excluded that the *Os* in Group C may have made use of a more refined verbal classification at the covert level.

In general, the problems of errors in a response system raises an interesting question concerning the number of different responses within a system that can be used effectively on a discrimination task. The present study shows that prediction of the stimuli would be increased if *O*'s second verbal response is used in the prediction in addition to the first. A possible explanation for this increased prediction is that the number of alternative responses in the first verbal system was insufficient for *O* to express all the discrimination he was capable of making. It is reasonable to assume that in some instances when an *O* in Group A gave as his first verbal judgment the response "8," he was uncertain as to whether the stimulus was actually "7" or "8." He had to decide between these two responses. In other instances when the response "8" was given he had had to decide between "8" and "9." The second verbal response thus gave *O* an opportunity to show the direction of his uncertainty.

A reasonable conclusion one might draw from this analysis is that the increased prediction accruing through the second verbal response would disappear if the number of alternative responses in the first verbal response system was increased. However, this conclusion does not fit with experimental observation. Eriksen and Hake (3) have shown on a very similar judgmental task that increasing the number of alternative responses in the system does not increase stimulus prediction over what is obtained when the number of different stimuli and the number of response alternatives are equal.

In terms of the model presented above, the increased prediction obtained with a second verbal response must result from noncorrelated errors in the two sets of verbal responses. One factor that might constitute at least part of this noncorrelated error is a difference in judgmental task that confronts *O* at the time of the second verbal judgment. For the first judgment *O* must attempt to deal with a continuum of size and to place the stimulus in one of 11 categories along this continuum. However, at the time of the second judgment *O* is faced with essentially a two-category judgment. He has already located the stimulus along the size dimension, and his judgment now is in terms of whether the stimulus is apt to be larger or smaller than his first judgment. This difference in judgmental task may be one of the factors subsumed under the noncorrelated error term.

Summary

Observers were presented with a size-judgment task under the method of absolute judgment. Three experimental groups were used. Group A used the number series 1–11 for their verbal judgments and concurrently moved a lever a distance proportional to the size of the presented simulus. In addition, they were also required to give a second-guess verbal judgment to each stimulus. Group B was identical to Group A with the exception that the lever response was made without the benefit of visual monitoring. Group C was allowed only two verbal response categories and no second guesses, but the lever was moved under the same conditions as Group A.

The results showed that the first and second verbal responses and the movement response each had a significant correlation with the stimuli when either or both of the other response systems were held constant. These findings are interpreted in terms of noncorrelated errors in the different response systems, and some of the factors contributing to these noncorrelated errors are discussed.

References

1. Eriksen, C. W. Subception: fact or artifact? *Psychol. Rev.*, 1956, **63**, 74–80.
2. Eriksen, C. W. An experimental and theoretical analysis of subception. *Amer. J. Psychol.*, 1956 **69**, 625–634.

3. Eriksen, C. W., & Hake, H. W. Absolute judgments as a function of the stimulus range and number of stimulus and response categories. *J. exp. Psychol.*, 1955, **49**, 323–332.

4. Eriksen, C. W., & Wechsler, H. Some effects of experimentally induced anxiety upon discrimination behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 458–463.

5. Garner, W. R., Hake, H. W., & Eriksen, C. W. Operationism and the concept of perception. *Psychol. Rev.*, 1956, **63**, 149–159.

6. Lazarus, R. S., & McCleary, R. A. Autonomic discrimination without awareness: a study of subception. *Psychol. Rev.*, 1951, **58**, 113–122.

(Received April 9, 1956)

Journal of Experimental Psychology
Vol. 53, No. 5, 1957

SUPPLEMENTARY REPORT: RESPONSE AVAILABILITY IN HUMAN TRIAL-AND-ERROR LEARNING[1]

CLYDE E. NOBLE
Louisiana State University

In a previous experiment (1) the writer demonstrated the influence of response availability (N_R) upon human performance in a nonverbal trial-and-error learning situation (2). When N_R was varied from 4 to 10 (holding task length constant), there was a progressive increase in the number of trials to mastery. The present study involves an extension of N_R from 11 to 14.

Method.—All details of the experiment, other than the variable condition, were the same as those used previously (1). In the four N_R treatments, the ratios of the number of relevant to available responses were 4/11, 4/12, 4/13, and 4/14. The incorrect keys ($R-$) were:

N_R	*Number of R− Key*
11	1-3--6---10------14----17--19
12	-2—5-7-9--11--13--15----18--
13	1-3-5-7--10----13--15--17--19
14	1-3-5-7-9--11--13--15--17--19

Each of the four N_R conditions ($n = 48$) was administered in a mixed factorial design under the 24 permutations of four reactions, with two *S*s appearing in each cell. The *S*s were 192 basic airmen at Lackland Air Force Base assigned randomly to the four treatments within each permutation.

Results.—The results are presented in Fig. 1 along with the data of the earlier study. The mean number of trials required to attain each criterion in terms of *first* choices (1R+) is plotted as a function of N_R. It appears that response availability has only a slight retarding effect upon the rate of trial-and-error learning under the present conditions.

Friedman's nonparametric analysis of variance test was applied to the distributions of ranks in mastery scores among the various conditions. For $df = 3$, $\chi^2 = 7.29$, which is only marginally significant ($.05 < P < .10$) of differences among the treatment effects. This result contrasts sharply with the highly significant differences ($P < .001$) obtained in the first experiment (1).

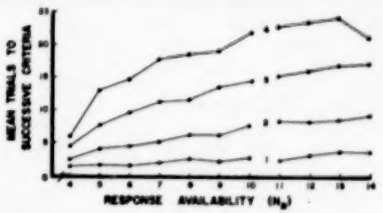


FIG. 1. Difficulty as a function of response availability (N_R), with the parameter indicating the four levels of proficiency. The curves are broken at $N_R = 10$ to denote the experimental independence of the two sets of data.

Inspection of response probability data revealed that each acquisition curve was sigmoidal and originated at a hypothetical value ($1/N_R$) representing the initial solubility of the problem. As before, there was no evidence of a serial position effect when errors were plotted over choice points.

[1] The experimental work was part of the program of the Air Force Personnel and Training Research Center. Airman 1/C J. B. Downs collected the data in 1952; the analysis of the data was supported by the Council on Research of Louisiana State University.

A $6 \times 4 \times 24$ analysis of variance was performed upon the 1R+ measure during the first six, five-trial blocks. By restricting the possible scores to the middle 80% range, extreme skewness and heterogeneity of form and variance were minimized. Practice is highly significant ($F = 566.21$; $df = 5/480$), but neither of the other two factors produces a main effect. The only interaction approaching significance is that between availability and permutations ($F = 1.19$; $df = 69/96$; $P = .10$).

The *initial* effects of the N_R treatments were tested by a 4×4 analysis of variance of correct first choices during Trials 1–4. From the evidence above, permutation differences as well as practice interactions may be assumed to be zero. The effect of trials is again highly significant ($F = 7.93$; $df = 3/564$) with no interaction, and the effect of availability is significant ($F = 3.73$; $df = 3/188$; $P < .025$). It appears that N_R produces real but small differences in performance early in training, but these effects disappear following the presolution period.

As may be observed in connection with Fig. 1, the possibility exists that some factor in the situation (e.g., "middleness") favors *Ss* having 14 available responses. To check this, the foregoing analysis was extended to Trial 5; by comparison, the practice variance becomes more important ($F = 28.51$; $df = 4/752$) and availability less so ($F = 1.98$; $df = 3/188$). Finally, the analysis was continued to include Trials 1–7; this resulted in the removal of all differences except those due to learning, which confirms the hypothesis that the correlation between difficulty and N_R is confined to the initial practice period in this experiment.

Discussion.—The absence of marked treatment effects supports the hypothesis of a decreasing rate of gain in difficulty with increasing behavior freedom. One could object, however, that the comparison of differences in the first experiment (1) with those found in this one is invalid. It might be argued that a χ^2 derived from a 7×24 matrix cannot be evaluated on the same basis as a χ^2 coming from a 4×24 matrix. In order to equalize the degrees of freedom, therefore, significance ratios were calculated separately for N_R varying from 4 to 7 and 7 to 10. For $df = 3$, the χ^2 values for the two sets of ranks are 30.46 and 8.36 respectively. The first is more significant ($P < .001$) than the second ($.02 < P < .05$), which in turn is more significant than the χ^2 of 7.29 ($.05 < P < .10$) obtained from the present study. These inequalities are quite in keeping with the hypothesis of a trend toward smaller differences with increases in N_R.

Considering both experiments, therefore, it is concluded that in human trial-and-error learning a negatively accelerated relationship exists between difficulty and N_R, and that the amount of over-all curvature varies with degree of proficiency.

References

1. Noble, C. E. Compound trial-and-error learning as a function of response availability (N_R). *J. exp. Psychol.*, 1955, **49**, 93–96.
2. Noble, C. E., & Farese, F. J. An apparatus for research in human selective learning. *J. Psychol.*, 1955, **39**, 475–484.

(Received July 7, 1956)